

**Swope, Sheridan**

---

**From:** Swope, Sheridan  
**Sent:** Wednesday, January 26, 2005 1:34 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** 10/649,273

For 10/649,273, pls search and interference search:

SID 2, full-length, against the NT and AA data bases  
SID 2, residues 148-414, against the NT and AA data bases  
SID 2, residues 176-414, against the NT and AA data bases

Sheridan Swope, Ph.D.  
Patent Examiner, AU 1652  
Recombinant Enzymes  
571-272-0943 (voice)  
E02B71 Remsen Bld (Office)  
E02C70 Remsen Bld (Mailbox)

10/649,273



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 143413**

**TO: Sheridan Swope**  
**Location: REM-2B71/2C70**  
**Art Unit: 1652**  
**Thursday, February 17, 2005**

**Case Serial Number: 10/649273**

**From: Mary Jane Ruhl**  
**Location: Biotech-Chem Library**  
**Remsen 1-A-62**  
**Phone: 571-272-2524**

**maryjane.ruhl@uspto.gov**

### **Search Notes**

Examiner Swope,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl  
Technical Information Specialist  
STIC  
Remsen 1-A-62  
Ext. 22524

From: Swope, Sheridan  
Sent: Wednesday, January 26, 2005 1:34 PM  
To: STIC-Biotech/ChemLib  
Subject: 10/649,273

For 10/649,273, pls search and interference search:

SID 2, full-length, against the NT and AA data bases  
SID 2, residues 148-414, against the NT and AA data bases  
SID 2, residues 176-414, against the NT and AA data bases

Sheridan Swope, Ph.D.  
Patent Examiner, AU 1652  
Recombinant Enzymes  
571-272-0943 (voice)  
E02B71 Remsen Bld (Office)  
E02C70 Remsen Bld (Mailbox)

\*\*\*\*\*

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2- \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA Sequence: # \_\_\_\_\_  
AA Sequence :# \_\_\_\_\_  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

## Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rnp**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

**When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.**



Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions .rnpm and .rnpn

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions .rapm and .rapn

***Because they contain data that is confidential, the results of Pending database searches should not be left in the case .***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 08:37:51 ; Search time 83.25 seconds  
(without alignments)  
1923.349 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125  
Sequence: 1 MLILRTAGVFPKRSRRKV.....DISKEVEASIKVDLKKMEI 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	414	5	ABG96478 Novel hum
2	2125	100.0	414	5	ABR05481 Human O-s
3	2125	100.0	414	6	ABJ26654 Human pro
4	2088	98.3	414	6	AAE29234 Human gly
5	2088	98.3	414	6	ABG71161 Novel hum
6	2088	98.3	414	6	ABU09569 Human gly
7	1845	86.8	364	6	ADA54471 Human pro
8	1385	65.2	267	5	ABG96487 Human pro
9	714.5	33.6	409	4	ABBB69133 Drosophila
10	659.5	31.0	439	3	AAAG19286 Arabidops
11	659.5	31.0	444	3	AAAG19285 Arabidops
12	634	29.8	463	3	AAV52216 Arabidops
13	549	25.8	179	5	ABG96489 Novel hum
14	524	24.7	382	6	ABU22934 Protein e
15	524	24.7	421	5	ABG96491 Novel hum
16	502	23.6	348	6	ABU35232 Protein e
17	501	23.6	350	8	ADL05040 M. catarr
18	499.5	23.5	401	7	ABO68626 Pseudomon
19	494.5	23.3	343	6	ABU39303 Protein e
20	492.5	23.2	341	4	AAU38205 Pseudomon
21	492.5	23.2	341	6	ABU38276 Protein e
22	492.5	23.2	341	7	ADG73342 P aerugin
23	490.5	23.0	341	7	ADG73344 P aerugin
24	489.5	23.0	335	6	ABU27480 Protein e
25	484.5	22.8	337	4	AAU34711 E. coli c

26	484.5	22.8	337	6	ABU47561 Protein e
27	484.5	22.8	337	6	ABU28771 Protein e
28	483.5	22.8	337	6	ABU50237 Protein e
29	482.5	22.7	340	6	ABU40514 Protein e
30	482.5	22.7	357	7	ADP06328 Bacterial
31	481.5	22.7	337	4	AAU38187 Salmonella
32	479.5	22.6	337	3	AAV52204 Escherich
33	475.5	22.4	325	3	AAV52203 Pasteurel
34	475.5	22.4	342	3	AAV52202 Hemophil
35	475.5	22.4	342	4	AAU35450 Hemophil
36	475.5	22.4	342	4	AAU35450 Hemophil
37	475.5	22.4	342	6	ABU30280 Protein e
38	474.5	22.3	343	7	ABO62704 Klebsiella
39	472.5	22.2	338	6	ABM67812 Phototrab
40	472	22.2	363	7	ADC96982 E. faeciu
41	469	22.1	340	6	ABU24130 Protein e
42	468	22.0	338	6	ABU29893 Protein e
43	465.5	21.9	346	3	AAV52208 Bacillus
44	465.5	21.9	354	6	ABU37844 Protein e
45	465	21.9	341	6	ABU40069 Protein e

## ALIGNMENTS

RESULT 1	ABG96478	standard; protein; 414 AA.
ID	ABG96478	
XX	ABG96478;	
AC		
XX		
DT	11-DEC-2002	(first entry)
XX		
DE	Novel human metalloprotease Mpi.	
XX		
KW	Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;	Abu47561 Protein e
KW	motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;	Abu28771 Protein e
KW	reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;	Abu50237 Protein e
KW	genital wart; metabolic disorder; premature puberty; Kallman syndrome;	Abu40514 Protein e
KW	Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;	Adp06328 Bacterial
KW	Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;	Abu38187 Salmonella
KW	liver disease; renal disease; immune disorder; rheumatoid arthritis;	AAV52204 Escherich
KW	acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;	AAV52203 Pasteurel
KW	emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;	AAV52202 Hemophil
KW	neurological disorder.	AAU35450 Hemophil
XX		
OS	Homo sapiens.	
XX		
PN	WO200272751-A2.	
XX		
PD	19-SEP-2002.	
XX		
PF	05-FEB-2002; 2002WO-US003353.	
XX		
PR	05-FEB-2001; 2001US-0266518P.	
XX		
PR	10-APR-2001; 2001US-0282814P.	
XX		
PA	(BRIM ) BRISTOL-MYERS SQUIBB CO.	
XX		
PI	Chen J, Feder J, Nelson TC, Duclos F, Krysstek S;	
XX		
DR	WPI; 2002-723329/78.	
XX		
DR	N-PSDB; ABS76635.	
XX		
PT	New isolated nucleic acid encoding MP-1 protein, useful for preventing,	Abu47561 Protein e
PT	treating, or ameliorating diseases associated with aberrant	Abu28771 Protein e
PT	metalloproteinase activity, e.g. immune, metabolic, inflammatory and	Abu50237 Protein e
PT	neurological disorders.	Abu40514 Protein e
XX		
PS	Claim 5; Fig 1A-C; 473pp; English.	Adp06328 Bacterial
XX		
CC	The invention describes an isolated nucleic acid molecule (I) encoding a	Abu38187 Salmonella
CC	metalloprotease (Mp-1). (I) is useful for preventing, treating, or	AAV52204 Escherich

CC ameliorating a medical condition, particularly an immune disorder, an  
 CC aberrant glutamate transport or motor neuron disorder, such as  
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
 CC condition. The compositions and methods are also useful for diagnosing,  
 CC prognosticating, treating, ameliorating and/or treating disorders  
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
 CC sequence of a metalloprotease Mpl protein

XX Sequence 414 AA;

Query Match 100.0%; Score 2125; DB 5; Length 414;

Best Local Similarity 100.0%; Pred. No. 2.2e-215; Mismatches 0; Indels 0; Gaps 0;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRKYEFELRSFNFHFGTLFLHKIVLGIEISCDPTAAAVDETSNV 60  
 DB 1 MLITKTAGVFFPKSKRKYEFELRSFNFHFGTLFLHKIVLGIEISCDPTAAAVDETSNV 60  
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEAISASGVSPSDLSAATITKPGI 120  
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEAISASGVSPSDLSAATITKPGI 120  
 QY 121 ALSIVGLSFSLSQVLGOLKKEPFIPIHMEAHALTITLTKNVEFPFLVLLISGHCILALV 180  
 DB 121 ALSIVGLSFSLSQVLGOLKKEPFIPIHMEAHALTITLTKNVEFPFLVLLISGHCILALV 180  
 QY 181 QGVSDFLILGKSLDIPGMDLKVARRSLIKHPECSWMSGKALIEHLAKOGRHFPIK 240  
 DB 181 QGVSDFLILGKSLDIPGMDLKVARRSLIKHPECSWMSGKALIEHLAKOGRHFPIK 240  
 QY 241 PPLHAKNCDFSPFTGLQHTYDKIIMKKEKEBIEGQILSSAADIAATVQHTMACHLVYR 300  
 DB 241 PPLHAKNCDFSPFTGLQHTYDKIIMKKEKEBIEGQILSSAADIAATVQHTMACHLVYR 300  
 QY 301 THRAILPCKORDLPPONNAVIVASGVASNFYIRRALETLTNAQTCTLLCPPRLCTDNG 360  
 DB 301 THRAILPCKORDLPPONNAVIVASGVASNFYIRRALETLTNAQTCTLLCPPRLCTDNG 360  
 QY 361 IMIANGIERLRAGIILHDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414  
 DB 361 IMIANGIERLRAGIILHDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

RESULT 2

ABJ05481 ID ABB05481 standard; protein; 414 AA.

XX ABB05481;

XX 19-APR-2002 (first entry)

XX Human O-6-ialoglycoproteinase-like protein SEQ ID NO:2.

XX Human; O-6-ialoglycoproteinase-like protein; OSGPLP; enzyme.

XX Homo sapiens.

XX CN1318550-A.

XX 24-OCT-2001.

XX 19-APR-2000; 2000CN-00106834.

XX 19-APR-2000; 2000CN-00106834.

XX

PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.

XX Mao Y, Xie Y;

XX WPI; 2002-115090/16.

XX N-PSDB; ABA93268.

PT O-6-ialoglycoproteinase-like protein and encoding polynucleotide, useful  
 for diagnosing, preventing and treating related diseases.

XX Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.

The present sequence represents human O-6-ialoglycoproteinase-like protein  
 (OSGPLP). The present invention also describes: (1) the preparation of  
 the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the  
 prevention and/or treatment of related diseases; (4) utilizing the OSGPLP  
 protein in screening its agonist, excitomotor and inhibitor and preparing  
 an antibody against the OSGPLP protein; and (5) the use of the OSGPLP  
 CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors  
 and antibodies in treating diseases related to the abnormal OSGPLP gene  
 CC and in preparing the medicine composite for the treatment

XX Sequence 414 AA;

Query Match 100.0%; Score 2125; DB 5; Length 414;

Best Local Similarity 100.0%; Pred. No. 2.2e-215; Mismatches 0; Indels 0; Gaps 0;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRKYEFELRSFNFHFGTLFLHKIVLGIEISCDPTAAAVDETSNV 60  
 DB 1 MLITKTAGVFFPKSKRKYEFELRSFNFHFGTLFLHKIVLGIEISCDPTAAAVDETSNV 60  
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEAISASGVSPSDLSAATITKPGI 120  
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEAISASGVSPSDLSAATITKPGI 120  
 QY 121 ALSIVGLSFSLSQVLGOLKKEPFIPIHMEAHALTITLTKNVEFPFLVLLISGHCILALV 180  
 DB 121 ALSIVGLSFSLSQVLGOLKKEPFIPIHMEAHALTITLTKNVEFPFLVLLISGHCILALV 180  
 QY 181 QGVSDFLILGKSLDIPGMDLKVARRSLIKHPECSWMSGKALIEHLAKOGRHFPIK 240  
 DB 181 QGVSDFLILGKSLDIPGMDLKVARRSLIKHPECSWMSGKALIEHLAKOGRHFPIK 240  
 QY 241 PPLHAKNCDFSPFTGLQHTYDKIIMKKEKEBIEGQILSSAADIAATVQHTMACHLVYR 300  
 DB 241 PPLHAKNCDFSPFTGLQHTYDKIIMKKEKEBIEGQILSSAADIAATVQHTMACHLVYR 300  
 QY 301 THRAILPCKORDLPPONNAVIVASGVASNFYIRRALETLTNAQTCTLLCPPRLCTDNG 360  
 DB 301 THRAILPCKORDLPPONNAVIVASGVASNFYIRRALETLTNAQTCTLLCPPRLCTDNG 360  
 QY 361 IMIANGIERLRAGIILHDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414  
 DB 361 IMIANGIERLRAGIILHDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

RESULT 3

ABJ26654 ID ABJ26654 standard; protein; 414 AA.

XX ABJ26654;

XX 01-MAY-2003 (first entry)

XX Human protein modification + maintenance molecule protein SEQ ID No. 8.

XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;

XX cerebrioprotective; antiparkinsonian; nootropic; antiinflammatory;

XX antitumor; hepatotropic; gynaecological; antibacterial; virucide;

XX protozoicide; antiparasitic; cell proliferative diseases; PMOD;

XX protein modification and maintenance molecule; immunogenic fragment;

XX cancer; autoimmune; inflammatory disease; neurological disorder;

KW gastrointestinal; developmental; vesicle trafficking disorder; infection;  
 KM protein-protein interaction; drug-target interaction;  
 KM gene expression profile; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WC0200300844-A2.  
 XX  
 PD 03-JAN-2003.  
 XX  
 PF 18-JUN-2002; 2002MO-US019360.  
 XX  
 XX 22-JUN-2001; 2001US-0300508P.  
 XX 06-JUL-2001; 2001US-0303445P.  
 PR 13-JUL-2001; 2001US-0305405P.  
 PR 09-AUG-2001; 2001US-0311442P.  
 PR 24-AUG-2001; 2001US-0314821P.  
 PR 29-AUG-2001; 2001US-0315929P.  
 PR 03-MAY-2002; 2002US-0378205P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Gandhi AR, Kabie AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM,  
 PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DM, Lee EA, Yue H;  
 PI Forsythe J, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;  
 PI Thangaveju K, Gietzen KJ, Ding L, Baughn MR, Borowsky ML, Yao MG;  
 PI Malia NK, Mason PM, Gururajan R, Lee S, Beecha SD, Lee SY, Tran UK;  
 PI Elliott VS, Luo W, Sprague WM, Tang YT, Lu Y, Zedardjadian Y;  
 XX  
 DR WPI; 2003-184039/18.  
 DR N-PSDB; ABT23207.  
 XX  
 PT New isolated human PMOD polypeptide and polynucleotide, useful for  
 PT diagnosing, treating and preventing diseases or conditions associated  
 PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  
 PT infections.  
 XX  
 PS Claim 63; Page 182-183; 225pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide comprising: any of 28  
 CC sequences of 48-1256 amino acids; a natural amino acid sequence at least  
 CC 98% identical to the 28 amino acid sequences, 94% identical to a sequence  
 CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
 CC acids, or 97% identical to a sequence of 242 amino acids; all given in  
 CC the specification; or a biologically active or immunogenic fragment of  
 CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
 CC in diagnosing, treating and preventing diseases or conditions associated  
 CC with the decreased expression of protein modification and maintenance  
 CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
 CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
 CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
 CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
 CC endometriosis), developmental, vesicle trafficking disorders, and  
 CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
 CC useful in assessing the effects of exogenous compounds on the expression  
 CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
 CC fragments are useful in screening compounds for effectiveness as agonist  
 CC or antagonist of the polypeptides, or in altering the expression of the  
 CC target polynucleotide and compounds that specifically bind to or modulate  
 CC the activity of the polypeptide. The microarray is useful in monitoring  
 CC or measuring protein-protein interactions, drug-target interactions, and  
 CC gene expression profiles. This sequence represents a human PMOD protein  
 CC of the invention  
 XX  
 SO Sequence 414 AA:

Query Match 100.0%; Score 2125; DB 6; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-215;  
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MLITKTAGVFPFKSKRVVFLSPFHPTLPLHKIVLIGLRSQDPTAAAVDERTNV 60  
 1 MLITKTAGVFPFKSKRVVFLSPFHPTLPLHKIVLIGLRSQDPTAAAVDERTNV 60

61 LGEAHSQTEVHLKGTGIVPPAAQOHRENIORIVGEALSASGVSPSDLSAATTKPGI 120  
 61 LGEAHSQTEVHLKGTGIVPPAAQOHRENIORIVGEALSASGVSPSDLSAATTKPGI 120  
 DB  
 QY 121 ALSIGVLSFSLQLVGQLKKPPIPIHMEAAHATITRLTNKVEPFLVLLISGHCILALV 180  
 DB 121 ALSIGVLSFSLQLVGQLKKPPIPIHMEAAHATITRLTNKVEPFLVLLISGHCILALV 180  
 QY 181 QGVSPDLLGSKLDIAPGDMILKVARSLIRHPECTSWGGVAIHLAQGRFHPDIK 240  
 DB 181 QGVSPDLLGSKLDIAPGDMILKVARSLIRHPECTSWGGVAIHLAQGRFHPDIK 240  
 QY 241 PLHAHAKNDFPFTGLOHTYDKTIIMKKEBEGEIKQIILSSAADIAATVQHTACILVVR 300  
 DB 241 PLHAHAKNDFPFTGLOHTYDKTIIMKKEBEGEIKQIILSSAADIAATVQHTACILVVR 300  
 QY 301 THRATILFCORDDLLPQNNAVLVAAGGVANFYIRRALBITLNNATQCTLLCPPRLCTDNG 360  
 DB 301 THRATILFCORDDLLPQNNAVLVAAGGVANFYIRRALBITLNNATQCTLLCPPRLCTDNG 360  
 QY 361 IMIANNGIERLRAGGILHDIGIRPERKCPGVDSKVGESIRVPLKMEI 414  
 DB 361 IMIANNGIERLRAGGILHDIGIRPERKCPGVDSKVGESIRVPLKMEI 414

RESULT 4  
 AAE29234  
 ID AAE29234 standard; protein; 414 AA.  
 AC  
 AC AAE29234;  
 XX  
 DT 27-JAN-2003 (first entry)  
 XX  
 DE Human glycoprotease 28472 protein.  
 XX  
 KW Human; adenosine deaminase; seven transmembrane domain receptor; cancer;  
 KW 7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arrhythmia;  
 KW rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;  
 KW hypertension; ischaemic heart disease; obesity; myocardial infarction;  
 KW endothelial cell disorder; Grave's disease; psoriasis; brain disorder;  
 KW Parkinson's disease; Alzheimer's disease; haematopoietic disorder;  
 KW cerebral oedema; metabolic disorder; liver disorder; platelet disorder;  
 KW chromosome mapping; tissue typing; gene therapy; neuroprotective;  
 KW cytosolic; anorectic; cardiac; haemostatic.  
 XX  
 OS Homo sapiens.  
 XX  
 PH Key  
 PH Location/Qualifiers  
 FT 1..108  
 FT /note= "Non-transmembrane domain; N-terminal cytoplasmic  
 FT domain"  
 FT 38..369  
 FT /note= "Endopeptidase O-sialoglycoprotein hydrolase  
 FT metalloprotease zinc glycoprotease sialoglycoprotease  
 FT domain"  
 FT 109..132  
 FT /note= "Transmembrane domain"  
 FT 133..164  
 FT /note= "Non-transmembrane domain; non-cytoplasmic loop"  
 FT 138..152  
 FT /note= "Glycoprotease domain"  
 FT 165..189  
 FT /note= "Transmembrane domain"  
 FT 190..316  
 FT /note= "Non-transmembrane domain; cytoplasmic domain"  
 FT 317..333  
 FT /note= "Transmembrane domain"  
 FT 334..414  
 FT /note= "Non-transmembrane domain"  
 FT 374..414  
 FT /note= "Sialoglycoprotease type domain"

PN WO200274960-A2.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 08-NOV-2001; 2001WO-US051427.  
 XX  
 PR 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Leihey KR, Kapeller-Libermann R, Glucksmann M;  
 XX  
 DR WPI: 2002-759898/82.  
 DR N-PSDB; AAD46856.  
 XX  
 PT New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,  
 PT useful for diagnosing and treating cancer, immune, cardiovascular,  
 PT hematopoietic, brain, pain, metabolic, liver or platelet disorders, and  
 PT in pharmacogenomics.  
 XX  
 PS Claim 1; Fig 8; 178pp; English.  
 XX  
 CC The present invention relates to novel 38650, 28472, 5495, 65507, 81588  
 CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-  
 CC protease or seven transmembrane domain (7TM) receptor family members.  
 CC Sequences of the invention are useful in diagnosing and treating cancer  
 CC or aberrant cellular proliferation and/or differentiation (e.g. colon or  
 CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid  
 CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.  
 CC hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,  
 CC myocardial infarction, thrombus) including endothelial cell disorders  
 CC (e.g. psoriasis, Grave's disease), hematopoietic disorders, brain  
 CC disorders (e.g. cerebral edema, Parkinson's or Alzheimer's disease),  
 CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet  
 CC disorders. They are also useful in screening assays, predictive medicine  
 CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials  
 CC and pharmacogenetics) and prophylactic and therapeutic methods. The  
 CC nucleic acids may also be used in chromosome mapping, tissue typing and  
 CC forensic biology and as surrogate markers. Sequences of the invention are  
 CC also used in gene therapy. The present sequence is human glycoprotease  
 CC 28472 protein.  
 XX  
 SQ Sequence 414 AA:  
 Query Match 98.3%; Score 2088; DB 5; Length 414;  
 Best Local Similarity 98.3%; Pred. No.1.9e-211;  
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 361 IMIANGIERLPAAGILHDIGIRYEPKPLGVDISKEVBA5IKVPOLAKMEI 414  
 DB 361 IMIANGIERLPAAGILHDIGIRYEPKPLGVDISKEVBA5IKVPOLAKMEI 414  
 RESULT 5  
 ABG71161  
 ID ABG71161 standard; protein; 414 AA.  
 XX  
 AC ABG71161;  
 XX  
 DT 30-JAN-2003 (first entry)  
 XX  
 DE Novel human glycoprotease 28472.  
 XX  
 KW Cancer; aberrant cell proliferation; aberrant cell differentiation;  
 KW breast cancer; ovarian cancer; prostate cancer; colon cancer;  
 KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;  
 KW endothelial disorder; hematopoietic disorder; blood vessel disorder;  
 KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;  
 KW platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;  
 KW autoimmune disorder; hypertension; atherosclerosis; heart failure;  
 KW myocardial infarction; ischaemic heart disease; Crohn's disease;  
 KW Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;  
 KW cerebral ischemia; peripheral neuropathy; Alzheimer's disease;  
 KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 138..152  
 FT /label=Glycoprotease\_domain  
 XX  
 PN WO200277233-A2.  
 XX  
 PT 03-OCT-2002.  
 XX  
 PD 08-NOV-2001; 2001WO-US046724.  
 XX  
 PF 08-NOV-2000; 2000US-0246768P.  
 XX  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Leihey KR, Kapeller-Libermann R, Glucksmann M;  
 XX  
 DR WPI: 2003-029938/02.  
 DR N-PSDB; ABS57020.  
 XX  
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain  
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,  
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or  
 PT hypertension.  
 XX  
 PS Claim 4; Fig 8A-B; 178pp; English.  
 XX  
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and  
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The  
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine  
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding  
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise  
 CC sequences that encode a human seven transmembrane domain (7TM). The  
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide  
 CC sequences are useful for diagnosing, preventing or treating a subject  
 CC with or at risk of developing a disorder, e.g. cancer or aberrant  
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,  
 CC prostate, colon or lung cancer), immune disorders, heart disorders,  
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,  
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,  
 CC liver disorders or platelet disorders. These disorders include carcinoma,  
 CC sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,

CC hypertension, atherosclerosis, heart failure, myocardial infarction,  
 CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki  
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, periaxial  
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,  
 CC cachexia or diabetes. This is the amino acid sequence of the novel human  
 CC glycoprotease 28472  
 CC  
 XX  
 XX Sequence 414 AA;

Query Match 98.3%; Score 2088; DB 6; Length 414;  
 Best Local Similarity 98.3%; Pred. No. 1.9e-211;  
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRKRYEFLRSFNHFGTLFLHKIVLGIEGSCDDTAAAVDEGTNV 60  
 DB 1 MLITKTAGVFFPKSKRKRYEFLRSFNHFGTLFLHKIVLGIEGSCDDTAAAVDEGTNV 60  
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIVEALSASGVSPSDLAATTIKPGL 120  
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIVEALSASGVSPSDLAATTIKPGL 120  
 QY 121 ALSLGVLSFSLQVLVGQKKPFIPIHMEAHALTIRLTNKVEPFLVLLISGHCCLALV 180  
 DB 121 ALSLGVLSFSLQVLVGQKKPFIPIHMEAHALTIRLTNKVEPFLVLLISGHCCLALV 180  
 QY 181 QGVSDFLILGKSLDAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKOGRFFPDIK 240  
 DB 181 QGVSDFLILGKSLDAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKOGRFFPDIK 240  
 QY 241 PPIHAKNCDFSFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYR 300  
 DB 241 PPIHAKNCDFSFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYR 300  
 QY 301 THRAILFCQKRDLLPONNAVIVASGVASNTYIRRALILTNAQCTLLCPPLRCTDNG 360  
 DB 301 THRAILFCQKRDLLPONNAVIVASGVASNTYIRRALILTNAQCTLLCPPLRCTDNG 360  
 QY 361 IMIANGIERLRAGILHDIGIRYBPKCPGVDSKEVGEASIKVQLKKEI 414  
 DB 361 IMIANGIERLRAGILHDIGIRYBPKCPGVDSKEVGEASIKVQLKKEI 414

## RESULT 6

ABU09569 standard; protein; 414 AA.

AC ABU09569;

DT 08-JUL-2003 (first entry)

DE Human glycoprotease encoded by cDNA 28472.

XX Human; enzyme; cancer; aberrant cellular proliferation; differentiation;  
 KW immune disorders; heart disorder; brain disorder;  
 KW cardiovascular disorder; endothelial cell disorder; pain disorder;  
 KW haematopoietic disorder; blood vessel disorder; metabolic disorder;  
 KW liver disorder; platelet disorder; glycoprotease.  
 XX Homo sapiens.

PN US2003009017-A1.

PD 09-JAN-2003.

PF 08-NOV-2001; 2001US-00012140.

PR 08-NOV-2000; 2000US-0246768P.

PR 08-NOV-2000; 2000US-0246772P.

PR 15-NOV-2000; 2000US-0249185P.

XX (LEIB) LEIBY K R.  
 PA (KAPR) KAPPLER-LIBERMANN R.  
 PA (GLUC) GLUCKSMANN M A.

XX Leiby KR, Kapeller-libermann R, Glucksmann MA;  
 PI WPI; 2003-42888/40.  
 DR N-PSDB; ACA60867.

PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,  
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic  
 PT and liver disorders.

PS Claim 4; Fig 8; 90pp; English.

XX  
 XX The invention relates to an isolated 38650 (encoding adenosine  
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
 CC a sequence which is at least 60% identical to the six nucleic acids or  
 CC their open reading frames, fragments of at least 15 nucleotides,  
 CC naturally occurring variants, or a DNA insert of the plasmid deposited  
 CC with the American Type Culture Collection as Accession No. not defined in  
 CC the specification, which encodes the amino acid sequence). Also included  
 CC are a host cell containing the nucleic acids (used to produce the  
 CC proteins), the encoded proteins, an antibody that selectively binds to  
 CC the polypeptide, and identifying a compound that binds to/modulates the  
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
 CC methods are useful for diagnosing, treating cancer, aberrant cellular  
 CC proliferation and/or differentiation, immune disorders, heart disorders,  
 CC cardiovascular disorders including endothelial cell disorders,  
 CC haematopoietic disorders, blood vessel disorders, brain disorders, pain  
 CC and metabolic disorders, liver disorders and platelet disorders (many  
 CC examples of these disorders are given in the specification). The present  
 CC sequence is the human glycoprotease encoded by cDNA 28472  
 XX

SQ Sequence 414 AA;

Query Match 98.3%; Score 2088; DB 6; Length 414;  
 Best Local Similarity 98.3%; Pred. No. 1.9e-211;  
 Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRKRYEFLRSFNHFGTLFLHKIVLGIEGSCDDTAAAVDEGTNV 60  
 DB 1 MLITKTAGVFFPKSKRKRYEFLRSFNHFGTLFLHKIVLGIEGSCDDTAAAVDEGTNV 60  
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIVEALSASGVSPSDLAATTIKPGL 120  
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIVEALSASGVSPSDLAATTIKPGL 120  
 QY 121 ALSLGVLSFSLQVLVGQKKPFIPIHMEAHALTIRLTNKVEPFLVLLISGHCCLALV 180  
 DB 121 ALSLGVLSFSLQVLVGQKKPFIPIHMEAHALTIRLTNKVEPFLVLLISGHCCLALV 180  
 QY 181 QGVSDFLILGKSLDAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKOGRFFPDIK 240  
 DB 181 QGVSDFLILGKSLDAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKOGRFFPDIK 240  
 QY 241 PPIHAKNCDFSFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYR 300  
 DB 241 PPIHAKNCDFSFTGLQHTVDKIIMKKEKEGIEKQIILSSADIAATVQHTMACHLYR 300  
 QY 301 THRAILFCQKRDLLPONNAVIVASGVASNTYIRRALILTNAQCTLLCPPLRCTDNG 360  
 DB 301 THRAILFCQKRDLLPONNAVIVASGVASNTYIRRALILTNAQCTLLCPPLRCTDNG 360  
 QY 361 IMIANGIERLRAGILHDIGIRYBPKCPGVDSKEVGEASIKVQLKKEI 414  
 DB 361 IMIANGIERLRAGILHDIGIRYBPKCPGVDSKEVGEASIKVQLKKEI 414

## RESULT 7

ADA54471 standard; protein; 364 AA.

AC ADA54471;

XX 20-NOV-2003 (first entry)  
DT Human protein, SEQ ID 2039.  
DE  
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;  
XX inflammatory disease; osteoporosis; neurological disease.  
OS Homo sapiens.  
XX  
XX EPI293569-A2.  
PN  
XX 19-MAR-2003.  
PD  
XX 21-MAR-2002; 2002EP-00006586.  
PF  
XX 14-SEP-2001; 2001JP-00328381.  
PR 24-JAN-2002; 2002US-0350435P.  
XX  
XX (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuhio Y;  
XX  
XX WPI: 2003-395539/38.  
DR N-PSDB; ADA52832.  
XX  
XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
XX Claim 14; SEQ ID NO 2039; 205pp; English.  
PS  
XX The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
XX Sequence 364 AA;  
SQ  
Query Match 86.8%; Score 1845; DB 6; Length 364;  
Best Local Similarity 99.2%; Pred. No. 7.9e-186;  
Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
XX 1 MLITKTAGVFFKSKRVYEFLLSFNHPGTLFLHAKIVLGIETSCDDPTAAVVDGTGV 60  
XX |||||  
XX 1 MLITKTAGVFFKSKRVYEFLLSFNHPGTLFLHAKIVLGIETSCDDPTAAVVDGTGV 60  
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALASAGVSPDLSAIAITIKPGL 120  
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALASAGVSPDLSAIAITIKPGL 120  
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIVOEALASAGVSPDLSAIAITIKPGL 120  
QY 121 ALISGVELSLSLOVGVQKRPPIPHHMEAHALTRLTNKVFPPLVLLISGHCILALV 180  
DB 121 ALISGVELSLSLOVGVQKRPPIPHHMEAHALTRLTNKVFPPLVLLISGHCILALV 180  
QY 121 ALISGVELSLSLOVGVQKRPPIPHHMEAHALTRLTNKVFPPLVLLISGHCILALV 180  
DB 121 ALISGVELSLSLOVGVQKRPPIPHHMEAHALTRLTNKVFPPLVLLISGHCILALV 180  
QY 181 QGVSDPFLILGKSLDIPGMDKVARSLIKHPECSMSGGKALEHAKOGNRPFIK 240  
DB 181 QGVSDPFLILGKSLDIPGMDKVARSLIKHPECSMSGGKALEHAKOGNRPFIK 240  
QY 181 QGVSDPFLILGKSLDIPGMDKVARSLIKHPECSMSGGKALEHAKOGNRPFIK 240  
DB 181 QGVSDPFLILGKSLDIPGMDKVARSLIKHPECSMSGGKALEHAKOGNRPFIK 240  
QY 241 PPIHAAKNCFSFTGLQHTVDKIKMKKEBGEIKGQILSSNADIAATVQHTMACHLYR 300  
DB 241 PPIHAAKNCFSFTGLQHTVDKIKMKKEBGEIKGQILSSNADIAATVQHTMACHLYR 300  
QY 241 PPIHAAKNCFSFTGLQHTVDKIKMKKEBGEIKGQILSSNADIAATVQHTMACHLYR 300  
DB 241 PPIHAAKNCFSFTGLQHTVDKIKMKKEBGEIKGQILSSNADIAATVQHTMACHLYR 300  
QY 301 THRAILFCQKRDLLPONNAVIVASGVASNYIRALAILTNATQCTLLCPPRLCTDNG 360  
DB 301 THRAILFCQKRDLLPONNAVIVASGVASNYIRALAILTNATQCTLLCPPRLCTDNG 360  
QY 301 THRAILFCQKRDLLPONNAVIVASGVASNYIRALAILTNATQCTLLCPPRLCTDNG 360  
DB 301 THRAILFCQKRDLLPONNAVIVASGVASNYIRALAILTNATQCTLLCPPRLCTDNG 360  
QY 361 IMIA 364

DB 361 IMIA 364  
|||||  
RESULT 8  
ABG96487  
ID ABG96487 standard; protein; 267 AA.  
XX  
XX ABG96487;  
AC  
XX 11-DEC-2002 (first entry)  
DT  
XX  
XX Novel human metalloprotease MPI fragment #1.  
DE  
XX  
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
KW neurological disorder.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200272751-A2.  
PN  
XX  
XX 19-SEP-2002.  
PD  
XX  
XX 05-FEB-2002; 2002WO-US003353.  
PF  
XX  
XX 05-FEB-2001; 2001US-0266518P.  
PR 10-APR-2001; 2001US-0282814P.  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
PA  
XX  
XX Chen J, Feder J, Nelson TC, Duclos F, Kyrstek S;  
PI  
XX  
XX WPI: 2002-723329/78.  
DR N-PSDB; ABS76639.  
DR  
XX  
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.  
XX  
XX Claim 5, Page 29; 473pp; English.  
PS  
XX The invention describes an isolated nucleic acid molecule (I) encoding a  
CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC diseases (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
CC sequence of a metalloprotease MPI protein  
XX  
XX Sequence 267 AA;  
SQ  
Query Match 65.2%; Score 1385; DB 5; Length 267;  
Best Local Similarity 100.0%; Pred. No. 2.4e-137;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 MEAHALTRITNKYEPFPLVLLISGHCCLALVGVSDPFLLGKSLDIAFGMDLKVARR 207  
DB 1 MEAHALTRITNKYEPFPLVLLISGHCCLALVGVSDPFLLGKSLDIAFGMDLKVARR 60  
QY 208 LSLIKHEPCSTMSGSKAIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYDKIIMKK 267  
DB 61 LSLIKHEPCSTMSGSKAIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYDKIIMKK 120  
QY 268 EKEGIEKGQILSSADIAATVQHTMACHLVKRTTRAILFCQKRDLLPQNNAVLVASGV 327  
DB 121 EKEGIEKGQILSSADIAATVQHTMACHLVKRTTRAILFCQKRDLLPQNNAVLVASGV 180  
QY 328 ASNFYIRRAEILTNATQCTLLCPPLCTDNGIMIANGIERLRAGLGLHDIGIRYE 387  
DB 181 ASNFYIRRAEILTNATQCTLLCPPLCTDNGIMIANGIERLRAGLGLHDIGIRYE 240  
QY 388 PKCPLGVDISKVEGEASIKVPQIKMEI 414  
DB 241 PKCPLGVDISKVEGEASIKVPQIKMEI 267

RESULT 9  
ABB69133 standard; protein; 409 AA.  
XX ABB69133;  
AC ABB69133;  
XX 26-MAR-2002 (first entry)  
DT 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 34191.  
DE Drosophila melanogaster polypeptide SEQ ID NO 34191.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical.  
XX Drosophila melanogaster.  
OS Drosophila melanogaster.  
XX WO200171042-A2.  
PN 27-SEP-2001.  
PD 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
PF 23-MAR-2001; 2000US-0191637P.  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
PA Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI, 2001-656860/75.  
XX N-PSDB; ABL13236.  
DR N-PSDB; ABL13236.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
interactions.  
XX Disclosure; SEQ ID NO 34191; 21pp + Sequence listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 409 AA;  
XX

Query Match 33.6%; Score 714.5; DB 4; Length 409;

Best Local Similarity 42.0%; Pred. No. 4,4e-66;  
Matches 156; Conservative 60; Mismatches 140; Indels 15; Gaps 6;  
QY 39 VGIENSQCDTAAAVVDENGVGEAISHQTEVHLTKTGIVPPAAQOLHRENIQRIYQEA 98  
DB 27 VGIETSCDDTGIAVDITGRVANYLSQOEHTHYGGIIPRAODLHRALESAVQRC 86  
QY 99 LSASGVSPEDLSAATTIRPGALSLGVLSFSLQVGLKPKPIPIHMEALATRLT 158  
DB 87 MEAAQAKPQVTRAIATTTTGPLSLLVGRFARHLARRQKLLPVHMEALQARME 146  
QY 159 N--KVEPPLVLLISGHCCLALVGVSDPFLLGKSLDIAFGMDLKVARRSLIKHEC 216  
DB 147 HBEQIGYPRCLLASGHCQLVANGPGRLLTGLQTLDAFGAPKIGRRLHLIPEX 206  
QY 217 SMSGSKAIEHLAK-QGNRPHFDIKPPLHAKNCDPSFTGLQHTYDKIIMKKEGIEK 275  
DB 207 RLWNGRALEHAAQSLADPLAYEFPLPLAQORCNPSFAGIKNSPRAIRARRARTTP 266  
QY 276 GQILSSADIAATVQHTMACHLVKRTTRAILFC--KQRDLLPQNNAVLVASGVASNFYI 333  
DB 267 DGVISNYGDFCAGLRSVSRHLMRTQRAIEYCLPFRQLFGDPTPLVWSSGVANNDAI 326  
QY 334 RRALEILTNATQCTLLCPPLCTDNGIMIANGIERLRAGLGLHDIGIRYE 392  
DB 327 VANIEHLAAQYGCGRSPRSKRYCSDNGVMIAHMGVEQL-----LQDKASTRYD--- 377  
QY 393 GVDISKVEGEA 403  
DB 378 SIDQSGAGFA 388

RESULT 10  
AAG19286 standard; protein; 439 AA.  
XX AAG19286;  
AC AAG19286;  
XX 17-OCT-2000 (first entry)  
DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21030.  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21030.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
PN 06-SEP-2000.  
PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
PF 25-FEB-2000; 1999US-0121825P.  
XX 25-FEB-1999; 1999US-0121825P.  
PR 05-MAR-1999; 1999US-0123180P.  
XX 05-MAR-1999; 1999US-0123180P.  
PR 23-MAR-1999; 1999US-0123548P.  
XX 23-MAR-1999; 1999US-0125788P.  
PR 25-MAR-1999; 1999US-0126264P.  
XX 25-MAR-1999; 1999US-0126264P.  
PR 29-MAR-1999; 1999US-0126785P.  
XX 29-MAR-1999; 1999US-0127462P.  
PR 01-APR-1999; 1999US-0128234P.  
XX 01-APR-1999; 1999US-0128234P.  
PR 06-APR-1999; 1999US-0128714P.  
XX 06-APR-1999; 1999US-0129845P.  
PR 16-APR-1999; 1999US-0130077P.  
XX 16-APR-1999; 1999US-0130077P.  
PR 21-APR-1999; 1999US-0130449P.  
XX 21-APR-1999; 1999US-0130510P.  
PR 23-APR-1999; 1999US-0130891P.  
XX 23-APR-1999; 1999US-0130891P.  
PR 28-APR-1999; 1999US-0131435P.  
XX 28-APR-1999; 1999US-0131435P.  
PR 30-APR-1999; 1999US-0132048P.  
XX 30-APR-1999; 1999US-0132048P.  
PR 30-APR-1999; 1999US-0132407P.  
XX 30-APR-1999; 1999US-0132407P.  
PR 04-MAY-1999; 1999US-0132484P.  
XX 04-MAY-1999; 1999US-0132484P.  
PR 05-MAY-1999; 1999US-0132485P.  
XX 05-MAY-1999; 1999US-0132485P.



PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132483P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.

PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154799P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.



PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 23-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148664P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151338P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.

PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158032P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160714P.  
PR 21-OCT-1999; 99US-0160715P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 22-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 31.0%; Score 659.5; DB 3; Length 444;

Best Local Similarity 40.8%; Pred. No. 3.3e-60; Matches 148; Conservative 57; Mismatches 127; Indels 31; Gaps 6;

QY 38 IYLGITSCDDPAAAVNDETNVIGBAIHSOTBEVHLKGTGIVPAAOQLRENIQRIYOE 97  
DB 90 VVIGITSCDDPAAAVNDETNVIGBAIHSOTBEVHLKGTGIVPAAOQLRENIQRIYOE 149  
QY 98 ALSASGVSPDLSAATTKIPGIALSLGVLSQVLQVQLKKPFIPIHMEAHALTIRL 157  
DB 150 ALDKANLTREKDLISAVAVTIGPGLSLCLRGVARRARVAGNFSLPYGVHMEHNTLVARL 209  
QY 158 T-NKVEFPPLVILISGCHLALVQGVSPFLIGKSLDIAPGMDLKVARRLSLIKHPBC 216  
DB 210 VQGLSFPFPMALLISGHNILVLAHKLQGYTQIGTVDDAIGBAFDTAKTAKGLDMH--- 266  
QY 217 STMSGKALIEHLAKQGNRHFPIKPLHAKNDFSFTGLQHVTDKIIMKKEKEGIEKG 276  
DB 267 --RSGGPAVEBELBEDDASVKFNVPMKYNKDCNFSYAGIKTQVRLAIEAKE---IDAK 320  
QY 277 QILSSA-----ADIAATVOHTMACHLVKRTTRAILFCFKORDLLPQNNNAVLVAGGVA 328  
DB 321 CPVSSATNEDRRNRADIIASFQVAVLHLEKGERAIDALE---LEPSIKHNVISGVA 377  
QY 329 SNFYIRALEILTNATQCTLLCPPLCTDNGIMIAMNGIERLPAIGILHDIIGRIYEP 388  
DB 378 SNKYVRLRNINYYENKNLKLVCPPSPCLCTDNGVMVAMVTEGHEHRRVG-----RYDP 427  
QY 389 KCP 391  
DB 428 PPP 430  
RESULT 12  
AAV52216  
ID AAV52216 standard; protein; 463 AA.  
XX  
AC AAV52216;  
XX  
DT 09-FEB-2000 (first entry)

XX Arabidopsis thaliana yjld protein homologue.  
 DE yjld protein; essential; Gram positive; Gram negative; conserved; motif;  
 XX identification; antagonist; antibacterial; antibiotic; broad spectrum;  
 KM treatment; infection; resistance; drug target.  
 OS Arabidopsis thaliana.  
 XX  
 XX Key Location/Qualifiers  
 FH Region 86..96 /note= "yjld conserved motif 3"  
 FT Region 111..131 /note= "yjld conserved motif 4"  
 FT Region 152..198 /note= "yjld conserved motif 2"  
 FT Region 208..259 /note= "yjld conserved motif 1"  
 FT Region /note= "yjld conserved motif 1"  
 XX MO9954470-A2.  
 XX  
 XX 28-OCT-1999.  
 XX  
 XX 20-APR-1999; 99WO-EP002635.  
 XX  
 XX 22-APR-1998; 98GB-00008423.  
 XX  
 XX (GLAX ) GLAXO GROUP LTD.  
 XX  
 XX Arigoni F, Edgerton MD, Loferer H, Petesch MC;  
 PI WPI; 2000-013253/01.  
 DR  
 XX Novel bacterial polypeptides used to identify broad spectrum antibiotics.  
 PT  
 PS Claim 1; Fig 1; 55pp; English.  
 XX  
 XX Proteins AAY52202-Y52217 encompass a novel family of proteins designated  
 CC the yjld family, after the name given to the Escherichia coli family  
 CC member. These proteins are essential for the survival of both Gram  
 CC negative and Gram positive bacteria, although no function has as yet been  
 CC ascribed to these proteins. The yjld proteins, fragments of yjld proteins  
 CC (for example, fragments encompassing one or more conserved yjld motifs  
 CC such as AAY52218-Y52284) and nucleotides encoding them can be used to  
 CC identify antagonists and broad spectrum antibacterial compounds. These  
 CC antagonists and compounds can be used to treat a wide range of bacterial  
 CC infections. New antibiotics are urgently needed, as serious bacterial  
 CC infections and antibiotic resistant strains are becoming increasingly  
 CC prevalent. The proteins of the invention are essential proteins for  
 CC bacterial viability, and represent new targets for antibiotics  
 CC  
 XX Sequence 463 AA;  
 SQ  
 XX  
 XX Query Match 29.8%; Score 634; DB 3; Length 463;  
 XX Best Local Similarity 37.9%; Pred. No. 1.8e-57;  
 XX Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;  
 QY 38 IVIGIETSCDDTAADVDEGTGVLGEAHSQTEVHLKTGTIVPPAAQQLHRENIQRIQVE 97  
 DB 85 VWIGIERSCDDTAADVDEGTGVLGEAHSQTEVHLKTGTIVPPAAQQLHRENIQRIQVE 97  
 QY 98 ALASAGVSPSLAIAATTKKGLALSLGVLSFSLQVLQKPPRIPIHNEAHLVTRL 157  
 DB 142 ALDPAANTERKDLGVAATVIGPGLSLCLFVGVKARVAGNSPSLPIVGVHNEAHLVTRL 201  
 QY 158 T-NKVEPPIVLLISGHCILALVGVSDPLLIGKSLIAPQMDLKVARRSLIKHPEC 216  
 DB 202 VBOELSPFPMLLSGGHNLVLAKHLQVYQLGTTVDALGEAPDKTKAKGLDMH--- 258  
 QY 217 STMSGGAIEPLATCCMMHFDIKFELTAAKCUFSFVLAHVTDKIIMKKEKEGIBKG 276  
 DB 259 --RSGGFAVEELALEGDAKSKVKNVPMKRYHDCNYSVAGLKTQVLALEAKS----- 308

QY 277 QILSSADIAATVQHTMACHLVKTRTRAILPFCKQDPLFPONNAVLSGCVASNPYTRRA 336  
 DB 309 --IRNRADIAASFQRAVAVHLLEKCEKRAIDWALE--LEPSIMHWVLSGGVSNKTVRLR 363  
 QY 337 LEILNNAQCTLLCPPEPRCTDNGMIAMNGIERAPAGILHDIGIRYE----- 387  
 DB 364 LNNIVENKRLKVCPPPSLCTDNGVAVMTGLEHFRVG-----RIDPPPAETPE 413  
 QY 388 -----PKCPGLGVDSKEVGEA 403  
 DB 414 DYVYDLRPRWPLGEETAKGRSFA 436  
 RESULT 13  
 ABG96489  
 ID ABG96489 standard; protein; 179 AA.  
 XX  
 XX ABG96489;  
 AC  
 XX 11-DEC-2002 (first entry)  
 DT  
 XX Novel human metalloprotease Mpl fragment #3.  
 DE  
 XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 KW neurological disorder.  
 XX  
 XX Homo sapiens.  
 OS  
 XX MO200272751-A2.  
 PN  
 XX 19-SEP-2002.  
 PD  
 XX 05-FEB-2002; 2002MO-US003353.  
 XX  
 XX 05-FEB-2001; 2001US-0266518P.  
 PR 10-APR-2001; 2001US-0282814P.  
 XX  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PA  
 XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
 PI WPI; 2002-723329/78.  
 DR  
 XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
 PT treating, or ameliorating diseases associated with aberrant  
 PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
 PT neurological disorders.  
 XX  
 XX Disclosure; Page 50; 473pp; English.  
 PS  
 XX The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
 CC ameliorating a medical condition, particularly an immune disorder, an  
 CC aberrant glutamate transport or motor neuron disorder, such as  
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
 CC condition. The compositions and methods are also useful for diagnosing,  
 CC prognosticating, treating, ameliorating and/or treating disorders  
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.

CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
 CC sequence of a metalloprotease MPI protein  
 XX  
 S0 Sequence 179 AA;  
 Query Match 25.8%; Score 549; DB 5; Length 179;  
 Best Local Similarity 66.5%; Pred. No. 3.8e-49;  
 Matches 119; Conservative 0; Mismatches 0; Indels 60; Gaps 1;  
 QY 38 IVLGIEISCDPTAAVVDENGVNIGEAHSQTEVHLKTGQIVPPAQQHRENIORIVOE 97  
 |||||  
 DB 1 IVLGIEISCDPTAAVVDENGVNIGEAHSQTEVHLKTGQIVPPAQQHRENIORIVOE 60  
 QY 98 ALSASGVSFSDLSAIAATIKPGLALSLGVGSFSLQVLGQKKPPIP----- 144  
 |||||  
 DB 61 ALSASGVSFSDLSAIAATIKPGLALSLGVGSFSLQVLGQKKPPIPCCATTCATCATAT 120  
 QY 145 -----IHHENAHALTR 156  
 DB 121 GGAGGCTCATGCACTTACTATTAGGTGACCAATAAGTAGAATTTCIHHENAHALTR 179  
 RESULT 14  
 ABU22934 ID  
 ABU22934 standard; protein; 382 AA.  
 XX  
 ABU22934;  
 XX  
 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #8461.  
 XX  
 Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 Bordetella pertussis.  
 OS  
 WC200277183-A2.  
 XX  
 03-OCT-2002.  
 PD  
 21-MAR-2002; 2002WO-US009107.  
 PF  
 21-MAR-2001; 2001US-00815242.  
 XX  
 06-SEP-2001; 2001US-00948993.  
 PR  
 25-OCT-2001; 2001US-0342923P.  
 PR  
 08-FEB-2002; 2002US-00072851.  
 PR  
 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;  
 PI Wall J, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 WPI: 2003-029926/02.  
 DR  
 N-PSDB; ACA26804.  
 DR  
 New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 Claim 25; SEQ ID NO 50858; 1766bp; English.  
 PS  
 The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 S0 Sequence 382 AA;  
 Query Match 24.7%; Score 524; DB 6; Length 382;  
 Best Local Similarity 37.3%; Pred. No. 5.7e-46;  
 Matches 134; Conservative 63; Mismatches 136; Indels 26; Gaps 10;  
 QY 29 HFG--TLFLHK--IVLGIEISCDPTAAVVDENGVNIGEAHSQTEVHLKTGQIVPPA 83  
 |||||  
 DB 24 HPGPRTLNVNSAPMIITLGFESSCDEGVAAVCTERGLAHALHTQTAMHQEYGVPELA 83  
 QY 84 QGLHRENIORIVOEALNSAGVSPDLSAIAATIKPGLALSLGVGSFSLQVLGQKKPPI 143  
 |||||  
 DB 84 SRDHLRRVPLTRQVLAENGTLTADGAVAYTAGPGLAQLVGAASVAQALANSRALPAL 143  
 QY 144 PIHHEAHALTRITLNT-KVEPFLVLLISGCHIALTVOGSDPDLTGLSKSLDIAPGMDL 202  
 |||||  
 DB 144 GIHHEGHLISPLALPPEPFPFVALVSGHTQMLVGVGVYELGERTLDAAAEAPD 203  
 QY 203 KVARRLSLIKPECESTMSGKAIEHLAKQGNRPHFDIKPPLHAKKNDPSFTGLQ-HVTD 261  
 |||||  
 DB 204 KSAKLMGL-GYP-----GGPALRLAEGQDSRVDLPRLMHSGLDPSFSGLKTAVVLT 256  
 QY 262 KIIMKKKKEBGEIKQIISLADIAATVGHMTACHVKTHTRALIFCKQDRLPQNNAYL 321  
 |||||  
 DB 257 RV--RAATRDGGELEB--ODRADLAAATQAAVVEVLAARAIKRL---KQYGL-----RRL 304  
 QY 322 VASGGVANSFYIRRALIEILTNATOCCTLCPPPLCTDNGIMTAMNGIERLAGIILHD 380  
 |||||  
 DB 305 VVAGGVGANALIRAHLARALKFLRAEAYFPPLSLCTDNGMILFAAEKVKAGLADIRE 363  
 RESULT 15  
 ABG96491 ID  
 ABG96491 standard; protein; 421 AA.  
 XX  
 ABG96491;  
 AC  
 11-DEC-2002 (first entry)  
 DT  
 XX  
 DB Novel human metalloprotease associated protein #2.  
 XX  
 Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 KW neurological disorder.  
 KW

OS Caenorhabditis elegans.  
XX  
PN WO200272751-A2.  
XX  
PD 19-SEP-2002.  
XX  
PF 05-FEB-2002; 2002WO-US003353.  
XX  
PR 05-FEB-2001; 2001US-0266518P.  
PR 10-APR-2001; 2001US-0282814P.  
XX  
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
XX  
DR WPI; 2002-723329/78.  
XX  
PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
XX neurological disorders.  
XX  
PS Disclosure; Page 465-466; 473pp; English.  
XX  
CC The invention describes an isolated nucleic acid molecule (1) encoding a  
CC metalloproteinase (MP-1). (1) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
CC sequence of a metalloproteinase MP1 associated protein  
XX  
SQ Sequence 421 AA;  
  
Query Match 24.7%; Score 524; DB 5; Length 421;  
Best Local Similarity 33.4%; Pred. No. 6.7e-46;  
Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8;  
  
QY 39 VLGIETSCDDPTAAVVDCTGVLEGAITHSOTREVLKKTGIVPPAAQQLHRENIQIVGEA 98  
DB 25 VLGIETSCDDPTAAVVDCTGVLEGAITHSOTREVLKKTGIVPPAAQQLHRENIQIVGEA 98  
QY 99 LSASGVSPDLSAATATTIKQGLALSLGVLSFSLQVLQKPKFPIHHMEAHALTIRLT 158  
DB 85 LNDAGTSPKLDNAVAVTTPGLVALKEGISAIGPAKKHRLPLIPVHMGAAHLSILV 144  
QY 159 -NKYEPFVLVLSGGHCLALVGVSDPFLKLSLDIAPGMDLDKVARLSLHNPES 217  
DB 145 DSGVRPFPSAVLISGGHALISVADVEFKLYGQSVSGSPGECIDKVARQLDGL- GSEFD 203  
QY 218 TMGGKAIKELHAKGN---RHPIKRPILHAKNCDPSFTGLQHYTDKIIMKEKEBGIE 274  
DB 204 GIHVGAANVELLASASADGHLRYPIPLPNVPRKANNFPOIKGSYLNILRLKNSETSID 263  
QY 275 KGOILSSADIATVQTMACHLVKTRTHAILFCORDLIPONNAVLVASGVASNPYIR 334  
DB 264 -----IPDFCASLQNTVAHHSKLIHFPESLSGQEKLPKQ--LVIGGVANQYIF 313  
QY 335 RALEILTNATQCTLLCPPRLCTDNGIMIANNGIERLAGIILHIDIGIRYRP----- 388  
DB 314 GAISKLSAANHVTTIKVLLSLCTDNAEMIAVSGL-----LMLVRSSEAIMWRPNDIPDT 367  
QY 389 -----KCPPLGVNDISKVEGASIKVPQLKM 412

Db 368 IYAHASDICTDASSEI-----IDTPRKL 392

Search completed: February 16, 2005, 13:05:54  
Job time : 91.25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using bw model

Run on: February 16, 2005, 12:43:41 ; Search time 24.75 Seconds  
(without alignments)  
1248.675 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125  
Sequence: 1 MLILTKTAGVFFPKPSKRYV.....DISKVEGASIKVPOLKMEI 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pdp: \*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pdp: \*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pdp: \*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pdp: \*  
5: /cgn2\_6/ptodata/1/iaa/PTUS.COMB.pdp: \*  
6: /cgn2\_6/ptodata/1/iaa/backfilest.pdp: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	414	US-10-067-443-2	Sequence 2, Appli
2	2090.5	98.4	439	US-10-067-443-19	Sequence 19, Appl
3	1385	65.2	267	US-10-067-443-32	Sequence 22, Appl
4	634	29.8	463	US-10-067-443-3	Sequence 3, Appli
5	549	25.8	179	US-10-067-443-25	Sequence 25, Appl
6	524	24.7	421	US-10-067-443-4	Sequence 4, Appli
7	524	24.7	421	US-10-067-443-28	Sequence 28, Appl
8	501	23.6	350	US-09-540-236-2726	Sequence 2726, Ap
9	499.5	23.5	401	US-09-252-991A-17372	Sequence 17372, A
10	488	23.0	342	US-08-087-797-3	Sequence 3, Appli
11	482.5	22.7	357	US-09-543-681A-6513	Sequence 6513, Ap
12	475.5	22.3	325	US-08-087-797-2	Sequence 2, Appli
13	474.5	22.3	343	US-09-489-039A-9221	Sequence 9221, Ap
14	472	22.2	363	US-09-107-532A-6609	Sequence 6609, Ap
15	461	21.7	336	US-08-987-121A-4	Sequence 4, Appli
16	457	21.5	327	US-10-067-443-5	Sequence 5, Appli
17	457	21.5	335	US-08-961-083-52	Sequence 52, Appl
18	457	21.5	335	US-09-536-784-52	Sequence 52, Appl
19	455	21.4	336	US-09-066-512-2	Sequence 2, Appli
20	454	21.4	366	US-09-114-000C-4956	Sequence 4956, Ap
21	453	21.3	336	US-09-583-110-4857	Sequence 4857, Ap
22	448	21.1	368	US-09-107-433-4221	Sequence 4221, Ap
23	443	20.8	344	US-09-134-001C-3909	Sequence 3909, Ap
24	433	20.1	341	US-03-145-624-2	Sequence 2, Appli
25	406.5	19.1	344	US-09-198-452A-213	Sequence 213, App
26	406.5	19.1	360	US-09-438-185A-196	Sequence 196, App
27	401.5	18.9	344	US-09-602-777A-148	Sequence 148, App

28	362	17.0	292	4	US-09-724-623-81	Sequence 81, Appl
29	359	16.9	214	4	US-09-328-352-4609	Sequence 4609, Ap
30	338.5	15.9	340	4	US-10-067-443-6	Sequence 6, Appli
31	319	15.0	273	4	US-09-710-279-728	Sequence 728, App
32	204	9.6	187	4	US-09-248-796A-19489	Sequence 19489, A
33	181.5	8.5	143	4	US-09-328-352-4387	Sequence 4387, Ap
34	109	5.1	515	4	US-09-252-991A-33046	Sequence 33046, A
35	100	4.7	225	4	US-09-602-777A-144	Sequence 144, App
36	97.5	4.6	580	4	US-09-489-039A-12013	Sequence 12013, A
37	97	4.6	601	3	US-09-155-036-5	Sequence 5, Appli
38	97	4.6	601	4	US-09-866-307-5	Sequence 5, Appli
39	97	4.6	614	3	US-09-155-036-6	Sequence 6, Appli
40	97	4.6	614	4	US-09-866-307-6	Sequence 6, Appli
41	96	4.5	612	4	US-09-802-540-13545	Sequence 13545, A
42	94.5	4.4	480	4	US-09-583-110-5050	Sequence 5050, Ap
43	94.5	4.4	481	4	US-09-107-433-3197	Sequence 3197, Ap
44	93.5	4.4	294	4	US-09-107-532A-6933	Sequence 6933, Ap
45	93.5	4.4	579	4	US-09-543-681A-5019	Sequence 5019, Ap

ALIGNMENTS

RESULT 1									
US-10-067-443-2									
; Sequence 2, Application US/10067443									
; Patent No. 6642041									
; GENERAL INFORMATION:									
; APPLICANT: Bristol-Myers Squibb Company									
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN									
; FILE REFERENCE: D0073 NP									
; CURRENT APPLICATION NUMBER: US/10/067,443									
; CURRENT FILING DATE: 2002-02-05									
; PRIOR APPLICATION NUMBER: US 60/266,518									
; PRIOR FILING DATE: 2001-02-05									
; PRIOR APPLICATION NUMBER: US 60/282,814									
; PRIOR FILING DATE: 2001-04-10									
; NUMBER OF SEQ ID NOS: 71									
; SOFTWARE: Patentin version 3.0									
; SEQ ID NO 2									
; LENGTH: 414									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-10-067-443-2									
Query Match 100.0%; Score 2125; DB 4; Length 414;									
Best Local Similarity 100.0%; Pred. No. 1.4e-234;									
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MLILTKTAGVFFPKPSKRYEFLRSFNFHGTLPFLHKIVLGIETSCDDPAAVVDENGV	60						
DB	1	MLILTKTAGVFFPKPSKRYEFLRSFNFHGTLPFLHKIVLGIETSCDDPAAVVDENGV	60						
QY	61	IGBAHSQREVLTKTGIVPPAAQOHLRENIQRIQVBAISASGVSSDSALATTKRGL	120						
DB	61	IGBAHSQREVLTKTGIVPPAAQOHLRENIQRIQVBAISASGVSSDSALATTKRGL	120						
QY	121	ALSLGVGSFSLQVLGOLKKEPIPIHMEAHALTITLTKVPEPPLVLIISGCHLALV	180						
DB	121	ALSLGVGSFSLQVLGOLKKEPIPIHMEAHALTITLTKVPEPPLVLIISGCHLALV	180						
QY	181	QGVSPFLILGKSLDIAFGDMLDKVARRSLIKHPECSTVSGGVAIEHLAKQGRFFFDIK	240						
DB	181	QGVSPFLILGKSLDIAFGDMLDKVARRSLIKHPECSTVSGGVAIEHLAKQGRFFFDIK	240						
QY	241	PLIHAACNDPSFTGLQHTDKIIMKKKEBGIKQQLISSAADIAATVQHTMACHLVGR	300						
DB	241	PLIHAACNDPSFTGLQHTDKIIMKKKEBGIKQQLISSAADIAATVQHTMACHLVGR	300						
QY	301	THRALTPCKQRDLFQNNNAVIVASGVASNPYIRALLETITNAQTOCTLLCPPRLLCTDNG	360						
DB	301	THRALTPCKQRDLFQNNNAVIVASGVASNPYIRALLETITNAQTOCTLLCPPRLLCTDNG	360						

QY 361 IMIANGIERLRAGILIHDIIEGIRYBKCPLGVDISKEVGEASIKVPOLKMEI 414  
 Db 361 IMIANGIERLRAGILIHDIIEGIRYBKCPLGVDISKEVGEASIKVPOLKMEI 414

## RESULT 2

US-10-067-443-19  
 ; Sequence 19, Application US/10067443  
 ; Patent No. 6642041

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED

TITLE OF INVENTION: SPINAL CORD, MP-1

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.0

SEQ ID NO 19

LENGTH: 439

TYPE: PR

ORGANISM: homo sapiens

US-10-067-443-19

Query Match

Best Local Similarity 93.8%; Score 2090.5; DB 4; Length 439;

Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVFPKSKRKYEFILRSFNHFGTLFLHAKIVLGIETSCDDTAAAVDETGNV 60  
 Db 1 MLITKTAGVFPKSKRKYEFILRSFNHFGTLFLHAKIVLGIETSCDDTAAAVDETGNV 60  
 QY 61 LGEAHSOTVHLKTGGIVPPAAQOLHRENTORIVQELASGVSPPSLAATIKRGL 120  
 Db 61 LGEAHSOTVHLKTGGIVPPAAQOLHRENTORIVQELASGVSPPSLAATIKRGL 120  
 QY 121 ALSIGVLSFSLQVGLKKEPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 180  
 Db 121 ALSIGVLSFSLQVGLKKEPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 180  
 QY 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSWSSGKALEHLAKQGNRHFPIK 240  
 Db 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSWSSGKALEHLAKQGNRHFPIK 240  
 QY 241 PPLHAKNCPESFTGLQHVTDKIMKEKEKEGI-----EK 275  
 Db 241 PPLHAKNCPESFTGLQHVTDKIMKEKEKEKEGI-----EK 275  
 QY 276 GQILSSAADIAATVQHTMACHLVKTTHRAILFCQKRDILLPNNNAVIVASGVASNFYIR 335  
 Db 276 GQILSSAADIAATVQHTMACHLVKTTHRAILFCQKRDILLPNNNAVIVASGVASNFYIR 335  
 QY 301 GQILSSAADIAATVQHTMACHLVKTTHRAILFCQKRDILLPNNNAVIVASGVASNFYIR 360  
 Db 301 GQILSSAADIAATVQHTMACHLVKTTHRAILFCQKRDILLPNNNAVIVASGVASNFYIR 360  
 QY 336 ALBILINATQCTLLCPRRLCTDNGIMIAMNGIEBLRAGLGIHDIIEGIRYBKCPLGVD 395  
 Db 336 ALBILINATQCTLLCPRRLCTDNGIMIAMNGIEBLRAGLGIHDIIEGIRYBKCPLGVD 395  
 QY 361 ALBILINATQCTLLCPRRLCTDNGIMIAMNGIEBLRAGLGIHDIIEGIRYBKCPLGVD 420  
 Db 361 ALBILINATQCTLLCPRRLCTDNGIMIAMNGIEBLRAGLGIHDIIEGIRYBKCPLGVD 420  
 QY 396 ISKEVGEASIKVPOLKMEI 414  
 Db 396 ISKEVGEASIKVPOLKMEI 414  
 QY 421 ISKEVGEASIKVPOLKMEI 439  
 Db 421 ISKEVGEASIKVPOLKMEI 439

## RESULT 3

US-10-067-443-22  
 ; Sequence 22, Application US/10067443  
 ; Patent No. 6642041

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED

FILE REFERENCE: D0073 NP  
 ; CURRENT APPLICATION NUMBER: US/10/067,443  
 ; CURRENT FILING DATE: 2002-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/266,518  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/282,814  
 ; PRIOR FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 22  
 ; LENGTH: 267  
 ; TYPE: PR  
 ; ORGANISM: homo sapiens  
 ; US-10-067-443-22

## Query Match

Best Local Similarity 65.2%; Score 1385; DB 4; Length 267;  
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 MEAHALITRLTNKVEPPLVLLISGHCILALVQGVSDFLILGKSLDIAPGMDLKVARR 207  
 Db 1 MEAHALITRLTNKVEPPLVLLISGHCILALVQGVSDFLILGKSLDIAPGMDLKVARR 207  
 QY 208 ISLKHPECSWSSGKALEHLAKQGNRHFPIKPLHAKNCPESFTGLQHVTDKIMKK 267  
 Db 208 ISLKHPECSWSSGKALEHLAKQGNRHFPIKPLHAKNCPESFTGLQHVTDKIMKK 267  
 QY 268 EKEEGIEKQILSSAADIAATVQHTMACHLVKTTHRAILFCQKRDILLPNNNAVIVASGV 327  
 Db 268 EKEEGIEKQILSSAADIAATVQHTMACHLVKTTHRAILFCQKRDILLPNNNAVIVASGV 327  
 QY 328 ASNFYIRRALBILINATQCTLLCPRRLCTDNGIMIAMNGIERLRAGLGIHDIIEGIRY 387  
 Db 328 ASNFYIRRALBILINATQCTLLCPRRLCTDNGIMIAMNGIERLRAGLGIHDIIEGIRY 387  
 QY 388 PKCPLGVDISKEVGEASIKVPOLKMEI 414  
 Db 388 PKCPLGVDISKEVGEASIKVPOLKMEI 414  
 QY 414 PKCPLGVDISKEVGEASIKVPOLKMEI 439  
 Db 414 PKCPLGVDISKEVGEASIKVPOLKMEI 439

## RESULT 4

US-10-067-443-3  
 ; Sequence 3, Application US/10067443  
 ; Patent No. 6642041

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED

TITLE OF INVENTION: SPINAL CORD, MP-1

FILE REFERENCE: D0073 NP

CURRENT APPLICATION NUMBER: US/10/067,443

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.0

SEQ ID NO 3

LENGTH: 463

TYPE: PR

ORGANISM: Arabidopsis thaliana

## Query Match

Best Local Similarity 29.8%; Score 634; DB 4; Length 463;  
 Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;

QY 38 IVLGIETSCDDTAAAVDETGNVILGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYOE 97  
 Db 38 IVLGIETSCDDTAAAVDETGNVILGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYOE 97  
 QY 85 VVLGIETSCDDTAAAVDETGNVILGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYOE 141  
 Db 85 VVLGIETSCDDTAAAVDETGNVILGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYOE 141  
 QY 98 ALSAGVSPSDLSAATITTKPGLALSIGVLSFSLQVGLKKEPFIPIHMEAHALITRL 157  
 Db 98 ALSAGVSPSDLSAATITTKPGLALSIGVLSFSLQVGLKKEPFIPIHMEAHALITRL 157



Db	142	ALDKANLFEKOLSAVAVTIIGPGLSICLARKVARKARVAAGFSLPIGVNHHMEHALVARL	201
Qy	158	T-NKVEFPFVLIIIGGGHCLLVGVSPFLLIGKSLDIAPGMDLKVARRLSLIXHPEC	216
Db	202	VEOEISFPFMALLISGGHNLVLAHKQGYOTGLTVDDAIGRAFPKTAKMLGLDMH---	258
Qy	217	STMSSGKAIENHLAKONRPFHDIKPRLHAKNCDSPFGTGOHYTDTKIIMKCEKEBGEK	276
Db	259	--RSGCPRAVEELATSGDASVCFNNPMKHKDCNCFYAGUKTVQVRLATEKE-----	308
Qy	277	QILSSAADIATVQHTMACHLVKTRTHAILFCQRDRLPQNNAVLVAASGGVASNPYIRRA	336
Db	309	--IRNRADIASFQVAVLHLEKCEBRAIDWALE--LPSIKHWVVISGGVASNKTVRLR	363
Qy	337	LEILTNALQCTLLCEPPRLCTONGIMIANGBLERLPAGIGILHDIGIRYE-----	387
Db	364	LNNIVENKRLKLVCEPPSLCTINGVAVMTGLEHFRVG-----RYDPPPPATEPE	413
Qy	388	-----PKCPLGVDISKEVEEA	403
Db	414	DYVDIDRPRWPLGEEVAKGRSBA	436

```

RESULT 5
US-10-067-443-25
; Sequence 25 Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED
; TITLE OF INVENTION: SPINAL CORD, NP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 179
; TYPE: PRT
; ORGANISM: homo sapiens
; US-10-067-443-25

```

RESULT 6  
 US-10-067-443-4  
 ; Sequence 4, Application US/10067443  
 ; Patent No. 6642041  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bristol-Myers Squibb Company;  
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
 ; TITLE OF INVENTION: SPINAL CORD, MP-1  
 ; FILE REFERENCE: D0073 NP

[illegible]

```

: ORGANISM: Caenorhabditis elegans
US-10-057-443-29

Query Match      24.7%; Score 524; DB 4; Length 421;
Best local Similarity 33.4%; Pred. No. 6,38-51;
Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8

```

```

QY 39 VLGJETSICDPPRAAVBETGNVLGAHISOREVHLKTGGIVPAPAOQLHRENIRIOVEA 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 25 VLGJETSICDPPRAAVBETGNVLGAHISOREVHLKTGGIVPAPAOQLHRENIRIOVEA 84
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 99 LSAAGVSPDSISAIATTIKPGLALSLGVGLSPSLQVGLQKPPRIPIHMEBAHLTTIRLT 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 LINDGATSPBKDLDAVAVTVPGLVIALKEGISAALGPAKGRHLPIPIVHHMAHALSTILV 144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 -NKVEPPLVLLIJGGHGLALVQGVSPFLLLKGSLLDAPDMDDKVARRLSLKHPCS 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 145 DDSVRFPFSSAVLLSGGHALISVADVEKFKLYGQSVSSPBCJCDKVARQLGDE-GSEFD 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 TMSGKXIEHLAKGN--REFPDIKPEPLHAKNCDFSTGLQHYTDKIINKKEKEBIE 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 204 GIHVGAIVEILLASRASADGHLRYPIFLPNVKANANPQIQKSTYLLERLRKNSFISID 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 KGOILSSAADIATVQHTMACHLVKTRTHAILFCRKQDRLPONNAVILVAGSGVANSFYIR 334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 -----IPDFCASIQNTVAHISIKLHIFPESISEGKLPKQ--LVIGGVANQYTF 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 335 RALELLMAVQCTLLCPPLRCTDNGIMIANNGIIBRLAAGIGLIHDIIGRIYEP----- 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 GAISKLSAAHVNTTIKVLISLCTDIAEMIAVSGL-----LMLVRSBAIMWRPNDIIPDT 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 389 -----KCPILGVDISKEVGEASIRKVPOLKM 412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 IYAHARSDIGTDASSEI-----IDTPRRKJL 392
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 8
US-09-540-236-2726
; Sequence 2726, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
APPLICANT: Gary L. Bretton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATALANICA
FILE REFERENCE: 2709/2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 2726
LENGTH: 350
TYPE: PRN
ORGANISM: M.catalanalis
US-09-540-236-2726

Query Match      23.6%; Score 501; DB 4; Length 350;
Best Local Similarity 34.5%; Pred. No.2e-48;
Matches 126; Conservative 56; Mismatches 139; Indels 44; Gaps 8;

QY       39 VLGIGTSCDDTTAAAVDET---GNVLGEAIIHSQTVEHLTKTGIVPPPAOQLHRENTORI   94
           |||||:::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB        6 VLGETSCEDEGLAIYDSTWMNGRGGVLSQQVYSQINLHATGCGVVPELASRDHIKRIKLPL    65
           :::::::||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY       95 VOEALASAGVSPSDLASAITIKPGLALSISGVGSFSQLQVGOLKKPIRPHIHEAHALT     154
           :>::::~||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB        66 FNEMLDQNMTTKSDIDAVALTYKGPGELIGALMTGALLFGRTLLYYGLGVPAVGHMEGHILIA   125
           ::::::::::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY      155 IRLTV-NKEVPFPFLVTILSGHCCLLATVOGVDPLLAKSLDIAPGMIDLVARBLSLIK     212
           |:|::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB      126 PLLASDPSPSPFPCCLVASGGHTMLVRADGVGVQIIGESI.DDAVGBECFDFTAKWLKI-P    184
           |::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY      213 HPECSTMSSGKAIEHLAQGRNFHFDIRPLELHAANCDFSTGTLOHTDTKIMMCKKEEG     272
           :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB      185 YP-----GGNIETELAKGNPNHAYELRPWQH-KGLDFSFGMKTAIHNLIKOTPNMQS     237
           :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY      273 IEKGILLSSADIATVQHTMAACHLVKRTHRAILLECRORDLIIPONNAVTLVAAGVANSFY     332
           :::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
DB      238 DP-----ATRADIASFERYAVVDTLYKCKCYRALOMTGIRO-----LVVAAGVANQT      284
           :::||::||::||::||::||::||::||::||::||::||::||::||::||::||:

QY      333 IRRAALELTNACTCCLCPPRPLCTDNCGIMTAMNGIERLRAG-----L              375
```

```

Db      285 LRLRTLTTLAQIDASVYVAATBELCTDNGAMIAIAGFORLSRGSGDDLAVRCIFRMWTML 344
QY      376 GILHD 380
      ||:|
Db      345 GIEYD 349

RESULT 9
US-09-252-991A-17372
; Sequence 17372, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ. ID NOS: 33142
; SEQ. ID NO 17372
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17372

Query Match      23.5%; Score 499.5; DB 4; Length 401;
Best Local Similarity 34.4%; Pred. No. 3,8e-48;
Matches 133; Conservative 67; Mismatches 158; Indels 29; Gaps 10;

QY      13 KPSKRYKYEFLRSFNFHFGTLPLHKI-----VLGIETSCDDTAAAVDEGNVLGEAII 66
Db      31 KASDRKTLVWRKADYIARLRICQALMLKPMRVLGLETSCDEGTVALYDSERGLADALF 90
QY      67 SOTEVHLKTGGIIVPPAAQOOLHRENIQRIQVQALSSAGSVSPSDLSAIIATIKGIALSLGV 126
Db      91 SQIDILHRVGGVVEPLASRDHVKRMELPLINQVLDSEGCPTADIALAYTAPGLVGLALV 150
QY      127 GLSPSLQVQGLKKPFIPIHMEAHALTITLTK-VEFPLVLLISGCHLLAVQGVSD 185
Db      151 GASCAQMAAPAWGVPAGVGHMHMECHLAPMLEQPPRFVALLVSGHQLQVLRVVDIGR 210
QY      186 FLILGKSLDIAFGDMDLKVARRLSLIGHPCSTMTSGSKALEHLAGKGNRHPDIKPPLIH 245
Db      211 YQLGSSVDVDAAGAFKTKMLTGL-GYP-----GGPEIARLAKERTPPGFVFPDPMTD 263
QY      246 AKQCDPFPTGLQHTWDKIIMKKEKEBGIKQGLISSADIAATVQHTMACHLVKTRTHAI 305
Db      264 RPLGDLFPFSGIKTFTTLN-TWQRCVENAGDDSEQ---TFCDIALAFQTAVVEETLLIKCRAL 319
QY      306 LFCQKQDILLPONNAVLVASGVSANFYIRKALELITLNATQCTLLCPPRLCTDNGIMIAW 365
Db      320 ---KQTL--KN---LVLAGVSANQALRSGLKEMKLGEMKGVAFYARPRECTDNGAMIAV 371
QY      366 NGIERLAAAGLILHDIGIRYEPKCP 392
Db      372 AGCQRLLAG--OHGPAISVQPRWPM 395

RESULT 10
US-08-087-797-3
; Sequence 3, Application US/08087797
; Patent No. 5543312
; GENERAL INFORMATION:
; APPLICANT: Mellors, Alan
; APPLICANT: Lo, Reggie Y.C.
; APPLICANT: Abdullah, Khalid M.
; TITLE OF INVENTION: Pasteurella Haemolytica
; TITLE OF INVENTION: Glycoprotease

```



TELEFAX: 704 334 2014  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 325 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-087-797-2

Query Match 22.4%; Score 475.5; DB 1; Length 325;  
 Best Local Similarity 35.8%; Pred. No. 1.5e-45;  
 Matches 120; Conservative 47; Mismatches 149; Indels 19; Gaps 6;

QY 39 VLGIETSCDDTAADVDEGTAVLGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIVOEA 98  
 DB 3 ILGIETSCDEGTAVIYDEDKGVANQIYSQIDMADYGGVPELSRDIHRTKLPIDQA 62  
 QY 99 LSASGVSPSDLSAATITIKPGIALSLVGSLQVLQVQKKPPIPIHMEAHALTIPL- 157  
 DB 63 LKEANLQPSDIDG:AVTAGPGVGLVGLSTIASLAVAMNVPALGVHMEGHLLAPMLE 122  
 QY 158 TNKVEPFLVLLISGGHCLLAVGVSDPFLILGKSLDIAPGMDLKVARRLSLIHPCS 217  
 DB 123 ENAEPEFPVALLISGHTQLVKVGVQYELIGESIDDAAGEAFDKTGLGL-DYP--- 178  
 QY 218 TMSGKALIEHLAKGNRFHDIKPLHAKNCDPSFTGLQHTVDKIIMKKEEGIEKQ 277  
 DB 179 ---AGVAMSKLASGTPRRFPRTMDRPGIDFSFGKTFPAANTIKANINENGELDEQ 235  
 QY 278 ILSASADIATVQHTMACHLVKTRTALIFCKORDLLPQNNAVLVSAGVSNFYIRAL 337  
 DB 236 ---TKCIDIAHAFQAV-----VDTIILKCK-RLAEQGYRLVMAGVSNKQRLADL 284  
 QY 338 EILTNAVQCTLLCPPEPLCTDNGIMIANNGIERLR 372  
 DB 285 AEMMKIKGEVFPYRPCTDNGAMIAVGTFLRK 319

RESULT 13  
 US-09-489-9221  
 ; Sequence 9221, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; CURRENT APPLICATION NUMBER: US/09/489, 039A  
 ; PRIOR FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 9221  
 ; LENGTH: 343  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 ; US-09-489-039A-9221

Query Match 22.3%; Score 474.5; DB 4; Length 343;  
 Best Local Similarity 34.5%; Pred. No. 2.1e-45;  
 Matches 123; Conservative 51; Mismatches 152; Indels 31; Gaps 8;

QY 39 VLGIETSCDDTAADVDEGTAVLGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIVOEA 98  
 DB 9 VLGIETSCDEGTAVIYDQOGLANQLYSQVTLADYGGVPELSRDIHRTKLPIDQA 68  
 QY 99 LSASGVSPSDLSAATITIKPGIALSLVGSLQVLQVQKKPPIPIHMEAHALTIPL- 157  
 DB 69 LKEANLQPSDIDG:AVTAGPGVGLVGLSTIASLAVAMNVPALGVHMEGHLLAPMLE 128  
 QY 158 TNKVEPFLVLLISGGHCLLAVGVSDPFLILGKSLDIAPGMDLKVARRLSLIHPCS 217  
 DB 129 DNPAFPFVALLVSGHTQLISVTGIGQYELIGESIDDAAGEAFDKTAKLGL-DYP--- 184

QY 218 TMSGKALIEHLAKGNRFHDIKPLHAKNCDPSFTGLQHTVDKIIMKKEEGIEKQ 277  
 DB 185 ---GGPMLSKASQTEGRFVFPRTMDRPGIDFSFGKTFPAANTIKANINENGELDEQ 235  
 QY 278 ILSASADIATVQHTMACHLVKTRTALIFCKORDLLPQNN- AVLVSAGVSNFYIRBA 336  
 DB 236 ---QTRADIARAFEDAVVDITLMIKCRRA-----LEQTFRLVMAGVSNRFTLRK 284  
 QY 337 EILTNAVQCTLLCPPEPLCTDNGIMIANNGIERLRPAIGIHLHDLE-GRIYKCPDL 392  
 DB 285 LAEMMQKRGVFPYRPCTDNGAMIAVGNVRLQTA----KALGVTRRWRPL 337

RESULT 14  
 US-09-107-532A-6609  
 ; Sequence 6609, Application US/09107532A  
 ; Patent No. 6563275  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 7310  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSER: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; CITY: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02354  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: PC  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,532A  
 ; FILING DATE: 30-Jun-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/085,598  
 ; FILING DATE: 14 May 1998  
 ; APPLICATION NUMBER: 60/051571  
 ; FILING DATE: July 2, 1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ariniello, Pamela Deneke  
 ; REGISTRATION NUMBER: 40,489  
 ; REFERENCE/DOCKET NUMBER: GTC-012  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (781)893-5007  
 ; TELEFAX: (781)893-8277  
 ; INFORMATION FOR SEQ ID NO: 6609:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 363 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEICAL: YES  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Enterococcus faecium  
 ; FEATURES:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (8) LOCATION 1...363  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6609:  
 ; US-09-107-532A-6609

Query Match 22.2%; Score 472; DB 4; Length 363;  
 Best Local Similarity 35.7%; Pred. No. 4.5e-45;  
 Matches 119; Conservative 59; Mismatches 123; Indels 32; Gaps 9;

QY 37 KIVGIETSCDDTAADVDEGTAVLGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIVO 96  
 DB 29 EILMIETSCDETSVAVVANGTEILSNIVASQINSHKRQGVVPEVASRHRHVQITLCLE 86

QY 97 EALSASGVSPDLSAIAATITKIPGALSLGVGLSFLQVLVGOLEKPEPIHNEAHALTIR 156  
 DB 89 DALVEAGASABDLAVAVTYGVLGSLILIGISAKAFAHQAQLPIPVNHAAGHLYAR 148  
 QY 157 LTNKVEPPLVLLISGHCILALVGVSDPFLIGKSLDIAPGMDLKVARRSLIKHPEC 216  
 DB 149 LVKFPPLMLLVSGGHTLVLMQEDGSYELIGETRDMAEADVKGRVIGL----- 202  
 QY 217 STMSGKAIEHLAKOG-NRPFEDIKPLHAKNCDFSTGLQHTVDKTIK-KEKEEGIE 274  
 DB 203 -SYSGKEIDDLAOGKDNHYF--PRAMIHEDNDFPSGSLKSATINLVHNAQORGEDID 259  
 QY 275 KGQILSSADIAATVQHTMACHLVKTRRAILFCQORDLLPQNNAV--LVASGVASNFY 332  
 DB 260 KN-----DLASFQASVIDVLTITLRA---C-----QNYVKGVLVAGVAAANG 302  
 QY 333 IRRALIELITNA--TQCTILCPPRLCTDNGIMI 363  
 DB 303 LREGLOALSAKLPEVELVIPPRLCGDNMAAMI 335

RESULT 15

US-08-987-121A-4  
 ; Sequence 4, Application US/08987121A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hoskins, Jo Ann  
 ; APPLICANT: Tang, Joseph Chiu-Chung  
 ; APPLICANT: Treadway, Patci Jean  
 ; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence  
 ; TITLE OF INVENTION: Gcp  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Eli Lilly and Company  
 ; STREET: Lilly Corporate Center  
 ; CITY: Indianapolis  
 ; STATE: Indiana  
 ; COUNTRY: U.S.  
 ; ZIP: 46285  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/987,121A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Webster, Thomas D.  
 ; REGISTRATION NUMBER: 39,872  
 ; REFERENCE/DOCKET NUMBER: X-  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 317-276-3334  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 336 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-987-121A-4

Query Match 21.7%; Score 461; DB 3; Length 336;  
 Best Local Similarity 34.5%; Pred. No. 7.3e-44;  
 Matches 116; Conservative 58; Mismatches 132; Indels 30; Gaps 8;

QY 37 KIVLGITSCDDPTAAVVDERTGNVLGRAIHSGTEVHLTKGIVPPAAQLRENIQRIYQ 96  
 DB 4 RYLLAFSTSCDEISVAALKNDLNSVIAQISHRKRGVPEVVASRHHVEVITACIB 63  
 QY 97 EALSAGCVSGDLSAIAATITKIPGALSLGVGLSFLQVLVGOLEKPEPIHNEAHALTIR 156  
 DB 64 EALABAGITBEDVAVVATYVPGVLGALLVGLSAKAPAMAHGCLPIPVNHAAGHLYAAQ 123

QY 157 LTNKVEPPLVLLISGHCILALVGVSDPFLIGKSLDIAPGMDLKVARRSLIKHPEC 216  
 DB 124 SVEPLEPPLIALLVSGGHTLVVVSAGDYKIIGETRDVAVGEAYDKVGMGL----- 177  
 QY 217 STMSGKAIEHLAKOGNRPFEDIKPLHAKNCDFSTGLQHTVDKTIKKEKEEGIEKG 276  
 DB 178 -TYPAGREIDELAHQCHDI-YDPPRAMIKEDNLEFSPGSKSAFINLHNAB-----QKG 230  
 QY 277 QILSSADIAATVQHTMACHLVKTRRAILFCQORDLLPQNNAVLVASGVASNFYIRRA 336  
 DB 231 ESLSLT-EDLCASFQAAVMDILMAKTKKAL-----EKYPVK--TLVAVGVAANKGLRER 281  
 QY 337 LBILTNATQCTILCPPRLCTDNGIMIA-----WN 366  
 DB 282 L--ATEITDVNVIIPRLCGDNAGMIAYASVSEWN 315

Search completed: February 16, 2005, 13:10:24  
 Job time : 27.75 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 13:08:47 ; Search time 239.85 seconds  
(without alignments)  
563.995 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 2125  
Sequence: 1 MLILTKTAGVFPKSRKRVY.....DISKVGESIKVPQLKMEI 414

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/BCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	414	US-10-067-443-2	Sequence 2, Appl1
2	2125	100.0	414	US-10-649-273-2	Sequence 2, Appl1
3	2125	100.0	414	US-10-651-722-2	Sequence 2, Appl1
4	2090.5	98.4	439	US-10-067-443-19	Sequence 19, Appl1
5	2090.5	98.4	439	US-10-649-273-19	Sequence 19, Appl1
6	2090.5	98.4	439	US-10-651-722-19	Sequence 19, Appl1
7	2088	98.3	414	US-10-012-140-5	Sequence 5, Appl1
8	1845	86.8	364	US-10-094-749-2039	Sequence 239, Ap
9	1385	65.2	267	US-10-067-443-22	Sequence 22, Appl
10	1385	65.2	267	US-10-649-273-22	Sequence 22, Appl
11	1385	65.2	267	US-10-651-722-22	Sequence 22, Appl
12	681.5	32.1	445	US-10-424-599-209259	Sequence 209259,
13	634	29.8	463	US-10-067-443-3	Sequence 3, Appl1

14	634	29.8	463	15	US-10-649-273-3	Sequence 3, Appl1
15	634	29.8	463	16	US-10-651-722-3	Sequence 3, Appl1
16	578	27.2	444	15	US-10-437-963-113732	Sequence 113732,
17	556.5	26.2	333	14	US-10-012-140-25	Sequence 25, Appl
18	549	25.8	179	14	US-10-067-443-25	Sequence 25, Appl
19	549	25.8	179	15	US-10-649-273-25	Sequence 25, Appl
20	549	25.8	179	15	US-10-651-722-25	Sequence 25, Appl
21	524	24.7	382	15	US-10-282-122A-50858	Sequence 50858, A
22	524	24.7	421	14	US-10-067-443-4	Sequence 4, Appl1
23	524	24.7	421	15	US-10-067-443-28	Sequence 28, Appl
24	524	24.7	421	15	US-10-649-273-4	Sequence 4, Appl1
25	524	24.7	421	15	US-10-649-273-28	Sequence 28, Appl
26	524	24.7	421	15	US-10-651-722-4	Sequence 4, Appl1
27	524	24.7	421	15	US-10-651-722-28	Sequence 28, Appl
28	511.5	24.1	347	14	US-10-012-140-24	Sequence 24, Appl
29	502	23.6	348	15	US-10-282-122A-63156	Sequence 63156, A
30	494.5	23.3	343	15	US-10-282-122A-67227	Sequence 67227, A
31	492.5	23.2	341	9	US-09-815-242-11798	Sequence 11798, A
32	492.5	23.2	341	15	US-10-282-122A-66200	Sequence 66200, A
33	489.5	23.0	335	15	US-10-282-122A-55404	Sequence 55404, A
34	484.5	22.8	337	15	US-09-815-242-10304	Sequence 10304, A
35	484.5	22.8	337	15	US-10-282-122A-56635	Sequence 56635, A
36	484.5	22.8	337	15	US-10-282-122A-75485	Sequence 75485, A
37	483.5	22.8	337	15	US-10-282-122A-78161	Sequence 78161, A
38	482.5	22.7	340	15	US-10-282-122A-68438	Sequence 68438, A
39	481.5	22.7	337	9	US-09-815-242-113780	Sequence 113780, A
40	475.5	22.4	342	9	US-09-815-242-11043	Sequence 11043, A
41	475.5	22.4	342	15	US-10-282-122A-58204	Sequence 58204, A
42	469	22.1	340	15	US-10-282-122A-52054	Sequence 52054, A
43	468	22.0	338	15	US-10-282-122A-57817	Sequence 57817, A
44	465.5	21.9	354	15	US-10-282-122A-65768	Sequence 65768, A
45	465	21.9	341	15	US-10-282-122A-67993	Sequence 67993, A

ALIGNMENTS

RESULT 1  
US-10-067-443-2  
; Sequence 2, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-067-443-2

Query Match 100.0%; Score 2125; DB 14; Length 414;  
Best Local Similarity 100.0%; Pred. No. 1.5e-199;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLILTKTAGVFPKSRKRVYFLRSFNHFGTLFLHKIVLGIETSCDDPAAVVDRTGV 60  
DB 1 MLILTKTAGVFPKSRKRVYFLRSFNHFGTLFLHKIVLGIETSCDDPAAVVDRTGV 60  
QY 61 LGEATNSQIVAVLNTAGIVFPAQQQLHRENTQRIQVQALNSAGVSPSDLSAATITIKPQL 120  
DB 61 LGEATNSQIVAVLNTAGIVFPAQQQLHRENTQRIQVQALNSAGVSPSDLSAATITIKPQL 120  
QY 121 ALSLGVGSFSLQVLGVQLKKPFIPIHMEAHALTRLTKVKEPPLVLLISGHCILALV 180

```
Db 121 ALSTGVGSLFSLQVLGQKKEPFIPIHMEBAHALTRITNKVEFPPLVLLISGHCILALV 180
|||
Qy 181 QGVSDPFLILGKSLDIPAGDMDKVARRSLIKHPECSTMSGKALIEHLAKQGNRRFPDIX 240
|||
Db 181 QGVSDPFLILGKSLDIPAGDMDKVARRSLIKHPECSTMSGKALIEHLAKQGNRRFPDIX 240
|||
Qy 241 PPLHAKKCDPFSFTGLQHTVDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Db 241 PPLHAKKCDPFSFTGLQHTVDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Qy 301 THRAILFCQKQDILPQNNNAVIVASGVASNFYIRRALBITLNAQTCTLLCPPRLCTDNG 360
|||
Db 301 THRAILFCQKQDILPQNNNAVIVASGVASNFYIRRALBITLNAQTCTLLCPPRLCTDNG 360
|||
Qy 361 IMIANGIERLRAGIGLIHDIEGIRYERKPCPLGVDSISKEVGASIKVPOLKMEI 414
|||
Db 361 IMIANGIERLRAGIGLIHDIEGIRYERKPCPLGVDSISKEVGASIKVPOLKMEI 414
|||

RESULT 2
US-10-649-273-2
; Sequence 2, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2

Query Match 100.0%; Score 2125; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.5e-199;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLILTKTAGVFFPKSRKRYEFLRSFNHPGTLFLHKIVLAGIETSCDDTPAAAVDDETVNV 60
|||
Db 1 MLILTKTAGVFFPKSRKRYEFLRSFNHPGTLFLHKIVLAGIETSCDDTPAAAVDDETVNV 60
|||
Qy 61 LGEAIIHSCQTEVHLKTKGIVPPAAQQLHRENIQRIVOEALISASGVSPSDLSAIAITTKPGI 120
|||
Db 61 LGEAIIHSCQTEVHLKTKGIVPPAAQQLHRENIQRIVOEALISASGVSPSDLSAIAITTKPGI 120
|||
Qy 121 ALSTGVGSLFSLQVLGQKKEPFIPIHMEBAHALTRITNKVEFPPLVLLISGHCILALV 180
|||
Db 121 ALSTGVGSLFSLQVLGQKKEPFIPIHMEBAHALTRITNKVEFPPLVLLISGHCILALV 180
|||
Qy 181 QGVSDPFLILGKSLDIPAGDMDKVARRSLIKHPECSTMSGKALIEHLAKQGNRRFPDIX 240
|||
Db 181 QGVSDPFLILGKSLDIPAGDMDKVARRSLIKHPECSTMSGKALIEHLAKQGNRRFPDIX 240
|||
Qy 241 PPLHAKKCDPFSFTGLQHTVDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Db 241 PPLHAKKCDPFSFTGLQHTVDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Qy 301 THRAILFCQKQDILPQNNNAVIVASGVASNFYIRRALBITLNAQTCTLLCPPRLCTDNG 360
|||
Db 301 THRAILFCQKQDILPQNNNAVIVASGVASNFYIRRALBITLNAQTCTLLCPPRLCTDNG 360
|||
Qy 361 IMIANGIERLRAGIGLIHDIEGIRYERKPCPLGVDSISKEVGASIKVPOLKMEI 414
|||
Db 361 IMIANGIERLRAGIGLIHDIEGIRYERKPCPLGVDSISKEVGASIKVPOLKMEI 414
|||
```

```
Db 361 IMIANGIERLRAGIGLIHDIEGIRYERKPCPLGVDSISKEVGASIKVPOLKMEI 414
|||

RESULT 3
US-10-651-722-2
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2

Query Match 100.0%; Score 2125; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 1.5e-199;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLILTKTAGVFFPKSRKRYEFLRSFNHPGTLFLHKIVLAGIETSCDDTPAAAVDDETVNV 60
|||
Db 1 MLILTKTAGVFFPKSRKRYEFLRSFNHPGTLFLHKIVLAGIETSCDDTPAAAVDDETVNV 60
|||
Qy 61 LGEAIIHSCQTEVHLKTKGIVPPAAQQLHRENIQRIVOEALISASGVSPSDLSAIAITTKPGI 120
|||
Db 61 LGEAIIHSCQTEVHLKTKGIVPPAAQQLHRENIQRIVOEALISASGVSPSDLSAIAITTKPGI 120
|||
Qy 121 ALSTGVGSLFSLQVLGQKKEPFIPIHMEBAHALTRITNKVEFPPLVLLISGHCILALV 180
|||
Db 121 ALSTGVGSLFSLQVLGQKKEPFIPIHMEBAHALTRITNKVEFPPLVLLISGHCILALV 180
|||
Qy 181 QGVSDPFLILGKSLDIPAGDMDKVARRSLIKHPECSTMSGKALIEHLAKQGNRRFPDIX 240
|||
Db 181 QGVSDPFLILGKSLDIPAGDMDKVARRSLIKHPECSTMSGKALIEHLAKQGNRRFPDIX 240
|||
Qy 241 PPLHAKKCDPFSFTGLQHTVDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Db 241 PPLHAKKCDPFSFTGLQHTVDKIKKKEKEBEGIEKQILSSAADIAATVQHTMACHLVVR 300
|||
Qy 301 THRAILFCQKQDILPQNNNAVIVASGVASNFYIRRALBITLNAQTCTLLCPPRLCTDNG 360
|||
Db 301 THRAILFCQKQDILPQNNNAVIVASGVASNFYIRRALBITLNAQTCTLLCPPRLCTDNG 360
|||
Qy 361 IMIANGIERLRAGIGLIHDIEGIRYERKPCPLGVDSISKEVGASIKVPOLKMEI 414
|||
Db 361 IMIANGIERLRAGIGLIHDIEGIRYERKPCPLGVDSISKEVGASIKVPOLKMEI 414
|||

RESULT 4
US-10-067-443-19
; Sequence 19, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
```

;; PRIOR FILING DATE: 2001-02-05  
;; PRIOR APPLICATION NUMBER: US 60/282,814  
;; PRIOR FILING DATE: 2001-04-10  
;; NUMBER OF SEQ ID NOS: 71  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 19  
;; LENGTH: 439  
;; TYPE: PRT  
;; ORGANISM: homo sapiens  
US-10-067-443-19

Query Match 98.4%; Score 2090.5; DB 14; Length 439;  
Best Local Similarity 93.8%; Pred. No. 4e-196;  
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVFFPKSKRYVEFLRSFNHFGTLPLHKIVLGIEISCDPTAAAVDSTGVN 60  
DB 1 MLITKTAGVFFPKSKRYVEFLRSFNHFGTLPLHKIVLGIEISCDPTAAAVDSTGVN 60  
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALSASGVSPDLSAIAITIKPGL 120  
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALSASGVSPDLSAIAITIKPGL 120  
QY 121 ALSIGVLSFSIQLVGQLKKPFIPIHMEAHALITRLTNKVEFPFLVLLISGHCILALV 180  
DB 121 ALSIGVLSFSIQLVGQLKKPFIPIHMEAHALITRLTNKVEFPFLVLLISGHCILALV 180  
QY 181 QGVSDPFLLGKSLDIPAGMDLKVARRSLIKHPECTMSGGKAIIEHLAKQGRHFDPDK 240  
DB 181 QGVSDPFLLGKSLDIPAGMDLKVARRSLIKHPECTMSGGKAIIEHLAKQGRHFDPDK 240  
QY 241 PPLHAKNCDPFSFTGLQHVTDKTIIMKKEKEBGI-----EK 275  
DB 241 PPLHAKNCDPFSFTGLQHVTDKTIIMKKEKEBGI-----EK 275  
QY 276 GQILSSADIAATVOHTMACHLVKRTTRAILFCQKRDLLPONNAVLVASGVASNFYIRR 335  
DB 301 GQILSSADIAATVOHTMACHLVKRTTRAILFCQKRDLLPONNAVLVASGVASNFYIRR 360  
QY 336 ALEITLTAATQCTLLCPPLCTDNGIMIAMNGIERLRAGLGIILHDEIGIRYEPKCPGLVD 395  
DB 361 ALEITLTAATQCTLLCPPLCTDNGIMIAMNGIERLRAGLGIILHDEIGIRYEPKCPGLVD 420  
QY 396 ISKEVGASIKVPOLKMEI 414  
DB 421 ISKEVGASIKVPOLKMEI 439

## RESULT 5

US-10-649-273-19  
;; Sequence 19, Application US/10649273  
;; Publication No. US20040043407A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Bristol-Myers Squibb Company  
;; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
;; FILE REFERENCE: D0073 CNT  
;; CURRENT APPLICATION NUMBER: US/10/649,273  
;; PRIOR FILING DATE: 2003-08-27  
;; PRIOR FILING DATE: 2001-02-05  
;; PRIOR FILING DATE: 2001-02-05  
;; PRIOR FILING DATE: 2002-02-05  
;; PRIOR FILING DATE: 2002-02-05  
;; PRIOR APPLICATION NUMBER: US 60/282,814  
;; PRIOR FILING DATE: 2001-04-10  
;; NUMBER OF SEQ ID NOS: 71  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 19  
;; LENGTH: 439  
;; TYPE: PRT  
;; ORGANISM: homo sapiens  
US-10-649-273-19

Query Match 98.4%; Score 2090.5; DB 15; Length 439;

Best Local Similarity 93.8%; Pred. No. 4e-196;  
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVFFPKSKRYVEFLRSFNHFGTLPLHKIVLGIEISCDPTAAAVDSTGVN 60  
DB 1 MLITKTAGVFFPKSKRYVEFLRSFNHFGTLPLHKIVLGIEISCDPTAAAVDSTGVN 60  
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALSASGVSPDLSAIAITIKPGL 120  
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALSASGVSPDLSAIAITIKPGL 120  
QY 121 ALSIGVLSFSIQLVGQLKKPFIPIHMEAHALITRLTNKVEFPFLVLLISGHCILALV 180  
DB 121 ALSIGVLSFSIQLVGQLKKPFIPIHMEAHALITRLTNKVEFPFLVLLISGHCILALV 180  
QY 181 QGVSDPFLLGKSLDIPAGMDLKVARRSLIKHPECTMSGGKAIIEHLAKQGRHFDPDK 240  
DB 181 QGVSDPFLLGKSLDIPAGMDLKVARRSLIKHPECTMSGGKAIIEHLAKQGRHFDPDK 240  
QY 241 PPLHAKNCDPFSFTGLQHVTDKTIIMKKEKEBGI-----EK 275  
DB 241 PPLHAKNCDPFSFTGLQHVTDKTIIMKKEKEBGI-----EK 275  
QY 276 GQILSSADIAATVOHTMACHLVKRTTRAILFCQKRDLLPONNAVLVASGVASNFYIRR 335  
DB 301 GQILSSADIAATVOHTMACHLVKRTTRAILFCQKRDLLPONNAVLVASGVASNFYIRR 360  
QY 336 ALEITLTAATQCTLLCPPLCTDNGIMIAMNGIERLRAGLGIILHDEIGIRYEPKCPGLVD 395  
DB 361 ALEITLTAATQCTLLCPPLCTDNGIMIAMNGIERLRAGLGIILHDEIGIRYEPKCPGLVD 420  
QY 396 ISKEVGASIKVPOLKMEI 414  
DB 421 ISKEVGASIKVPOLKMEI 439

## RESULT 6

US-10-651-722-19  
;; Sequence 19, Application US/10651722  
;; Publication No. US20040048302A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Bristol-Myers Squibb Company  
;; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
;; FILE REFERENCE: D0073 DIV  
;; CURRENT APPLICATION NUMBER: US/10/651,722  
;; PRIOR FILING DATE: 2003-08-29  
;; PRIOR FILING DATE: 2003-08-29  
;; PRIOR FILING DATE: 2001-02-05  
;; PRIOR FILING DATE: 2001-02-05  
;; PRIOR FILING DATE: 2002-02-05  
;; PRIOR FILING DATE: 2002-02-05  
;; PRIOR APPLICATION NUMBER: US 60/266,518  
;; PRIOR APPLICATION NUMBER: US 60/282,814  
;; PRIOR FILING DATE: 2001-04-10  
;; NUMBER OF SEQ ID NOS: 71  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 19  
;; LENGTH: 439  
;; TYPE: PRT  
;; ORGANISM: homo sapiens  
US-10-651-722-19

Query Match 98.4%; Score 2090.5; DB 15; Length 439;  
Best Local Similarity 93.8%; Pred. No. 4e-196;  
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLITKTAGVFFPKSKRYVEFLRSFNHFGTLPLHKIVLGIEISCDPTAAAVDSTGVN 60  
DB 1 MLITKTAGVFFPKSKRYVEFLRSFNHFGTLPLHKIVLGIEISCDPTAAAVDSTGVN 60  
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALSASGVSPDLSAIAITIKPGL 120  
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIYQEALSASGVSPDLSAIAITIKPGL 120  
QY 121 ALSIGVLSFSIQLVGQLKKPFIPIHMEAHALITRLTNKVEFPFLVLLISGHCILALV 180



```

Db 121 ALSTGVGVSFSLQVLGVQLKKPPIPIHMEAHALTRLTNRKVFPLVLLISGGHCLALV 180
Qy 181 QGVSDPFLLGKSLDIPAGDMDKVARRLSLIKHPECSMSGKALBHLAKQNRHFPHDIK 240
Db 181 QGVSDPFLLGKSLDIPAGDMDKVARRLSLIKHPECSMSGKALBHLAKQNRHFPHDIK 240
Qy 241 PPLHAKNCDSFPTGLQHTVDKTIIMKEKEBEGI-----EK 275
Db 241 PPLHAKNCDSFPTGLQHTVDKTIIMKEKEBEGIPLISKEVEQINIGLCLKIAHCFREK 300
Qy 276 GQILSSAADIAATVQHTMACHLVYKTRHAILFCQKRDLLPQNNAVLVASGVASNFYIR 335
Db 301 GQILSSAADIAATVQHTMACHLVYKTRHAILFCQKRDLLPQNNAVLVASGVASNFYIR 360
Qy 336 ALBILTAATQCTLLCPPRCLTDNGIMIANNGIERLRAGLGIHDIEGIRYEPKPLGYD 395
Db 361 ALBILTAATQCTLLCPPRCLTDNGIMIANNGIERLRAGLGIHDIEGIRYEPKPLGYD 420
Qy 396 ISKEVGASIKVPOLKMEI 414
Db 421 ISKEVGASIKVPOLKMEI 439

RESULT 7
US-10-012-140-5
; Sequence 5, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-140-5

Query Match 98.3%; Score 2088; DB 14; Length 414;
Best Local Similarity 98.3%; Pred. No. 6.4e-196;
Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MLILTKTAGVFPKSKRKRYEFLRSFNHPTGLFLHKLIVLGITSCDDTAAAVVDEGTAV 60
Db 1 MLILTKTAGVFPKSKRKRYEFLRSFNHPTGLFLHKLIVLGITSCDDTAAAVVDEGTAV 60
Qy 61 LGEAHSQTEVHLKTKGIVPPAAQOLHRENIQRIYQEALASAGVSPSDLSAATITIKPGL 120
Db 61 LGEAHSQTEVHLKTKGIVPPAAQOLHRENIQRIYQEALASAGVSPSDLSAATITIKPGL 120
Qy 121 ALSTGVGVSFSLQVLGVQLKKPPIPIHMEAHALTRLTNRKVFPLVLLISGGHCLALV 180
Db 121 ALSTGVGVSFSLQVLGVQLKKPPIPIHMEAHALTRLTNRKVFPLVLLISGGHCLALV 180
Qy 181 QGVSDPFLLGKSLDIPAGDMDKVARRLSLIKHPECSMSGKALBHLAKQNRHFPHDIK 240
Db 181 QGVSDPFLLGKSLDIPAGDMDKVARRLSLIKHPECSMSGKALBHLAKQNRHFPHDIK 240
Qy 241 PPLHAKNCDSFPTGLQHTVDKTIIMKEKEBEGIBKQILSSAADIAATVQHTMACHLVYR 300

```

```

Db 241 PPLHAKNCDSFPTGLQHTVDKTIIMKEKEBEGIBKQILSSAADIAATVQHTMACHLVYR 300
Qy 301 THRALIFCQKRDLLPQNNAVLVASGVASNFYIRALBILTAATQCTLLCPPRCLTDNG 360
Db 301 THRALIFCQKRDLLPQNNAVLVASGVASNFYIRALBILTAATQCTLLCPPRCLTDNG 360
Qy 361 IMIANNGIERLRAGLGIHDIEGIRYEPKPLGYDISKEVGASIKVPOLKMEI 414
Db 361 IMIANNGIERLRAGLGIHDIEGIRYEPKPLGYDISKEVGASIKVPOLKMEI 414

RESULT 8
US-10-094-749-2039
; Sequence 2039, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: MAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYUKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2039
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2039

Query Match 86.8%; Score 1845; DB 15; Length 364;
Best Local Similarity 99.2%; Pred. No. 3.8e-172;
Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MLILTKTAGVFPKSKRKRYEFLRSFNHPTGLFLHKLIVLGITSCDDTAAAVVDEGTAV 60
Db 1 MLILTKTAGVFPKSKRKRYEFLRSFNHPTGLFLHKLIVLGITSCDDTAAAVVDEGTAV 60
Qy 61 LGEAHSQTEVHLKTKGIVPPAAQOLHRENIQRIYQEALASAGVSPSDLSAATITIKPGL 120
Db 61 LGEAHSQTEVHLKTKGIVPPAAQOLHRENIQRIYQEALASAGVSPSDLSAATITIKPGL 120
Qy 121 ALSTGVGVSFSLQVLGVQLKKPPIPIHMEAHALTRLTNRKVFPLVLLISGGHCLALV 180
Db 121 ALSTGVGVSFSLQVLGVQLKKPPIPIHMEAHALTRLTNRKVFPLVLLISGGHCLALV 180
Qy 181 QGVSDPFLLGKSLDIPAGDMDKVARRLSLIKHPECSMSGKALBHLAKQNRHFPHDIK 240
Db 181 QGVSDPFLLGKSLDIPAGDMDKVARRLSLIKHPECSMSGKALBHLAKQNRHFPHDIK 240
Qy 241 PPLHAKNCDSFPTGLQHTVDKTIIMKEKEBEGIBKQILSSAADIAATVQHTMACHLVYR 300

```

Db 241 PRLHAKNCDPFTGLQHTVDKIIKKKEGIEKGQILSSADIAATVQHTMACHLVKR 300  
QY 301 THRAILFCQKORDLLPQNNNAVIVASGVSANFYIRALELTNAOCTLLCPPRLCTDNG 360  
Db 301 THRAILFCQKORDLLPQNNNAVIVASGVSANFYIRALELTNAOCTLLCPPRLCTDNG 360  
QY 361 IMIA 364  
Db 361 IMIA 364

## RESULT 9

US-10-667-443-22  
; Sequence 22, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-667-443-22

Query Match 65.2%; Score 1385; DB 14; Length 267;  
Best Local Similarity 100.0%; Pred. No. 3.5e-127;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 MEAALTLRLTNKVEFPPLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLVARR 207  
Db 1 MEAALTLRLTNKVEFPPLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLVARR 60  
QY 208 LSLIKHPECSMTSGGKAI EHLAKQGNRFPHDIKPLHAKNCDPFTGLQHTVDKIIKK 267  
Db 61 LSLIKHPECSMTSGGKAI EHLAKQGNRFPHDIKPLHAKNCDPFTGLQHTVDKIIKK 120  
QY 268 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 327  
Db 121 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 180  
QY 328 ASNFYIRALELTNAOCTLLCPPRLCTDNGIMIANGIERLRAGLGIHDIGIRYE 387  
Db 181 ASNFYIRALELTNAOCTLLCPPRLCTDNGIMIANGIERLRAGLGIHDIGIRYE 240  
QY 388 PKCPLGVDISKVEGASIKVPOLKMEI 414  
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

## RESULT 10

US-10-649-273-22  
; Sequence 22, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05

; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-649-273-22

Query Match 65.2%; Score 1385; DB 15; Length 267;  
Best Local Similarity 100.0%; Pred. No. 3.5e-127;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 MEAALTLRLTNKVEFPPLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLVARR 207  
Db 1 MEAALTLRLTNKVEFPPLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLVARR 60  
QY 208 LSLIKHPECSMTSGGKAI EHLAKQGNRFPHDIKPLHAKNCDPFTGLQHTVDKIIKK 267  
Db 61 LSLIKHPECSMTSGGKAI EHLAKQGNRFPHDIKPLHAKNCDPFTGLQHTVDKIIKK 120  
QY 268 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 327  
Db 121 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 180  
QY 328 ASNFYIRALELTNAOCTLLCPPRLCTDNGIMIANGIERLRAGLGIHDIGIRYE 387  
Db 181 ASNFYIRALELTNAOCTLLCPPRLCTDNGIMIANGIERLRAGLGIHDIGIRYE 240  
QY 388 PKCPLGVDISKVEGASIKVPOLKMEI 414  
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

## RESULT 11

US-10-651-722-22  
; Sequence 22, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-651-722-22

Query Match 65.2%; Score 1385; DB 15; Length 267;  
Best Local Similarity 100.0%; Pred. No. 3.5e-127;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 MEAALTLRLTNKVEFPPLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLVARR 207  
Db 1 MEAALTLRLTNKVEFPPLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLVARR 60  
QY 208 LSLIKHPECSMTSGGKAI EHLAKQGNRFPHDIKPLHAKNCDPFTGLQHTVDKIIKK 267  
Db 61 LSLIKHPECSMTSGGKAI EHLAKQGNRFPHDIKPLHAKNCDPFTGLQHTVDKIIKK 120  
QY 268 EKEGIEKGQILSSADIAATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVIVASGV 327

Db 121 EKEBIEKQIILSSAADIAATVQHTMACHLVYKTRRAILFCQKORDLLPONNAVLVASGV 180  
Qy 328 ASNFYIRALIELTNAQCTLLCPPEPLCTDNGIMIANNGIERLRAAGILHDIGIRYE 387  
Db 181 ASNFYIRALIELTNAQCTLLCPPEPLCTDNGIMIANNGIERLRAAGILHDIGIRYE 240  
Qy 388 PKCPLGVDISKVEGASIKVPOLKMEI 414  
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

## RESULT 12

US-10-424-599-209259  
; Sequence 209259, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 209259  
; LENGTH: 445  
; TYPE: PRP  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_WRT3847\_3098C.1.pep  
US-10-424-599-209259

Query Match 32.1%; Score 681.5; DB 15; Length 445;  
Best Local Similarity 43.1%; Pred. No. 8.8e-58;  
Matches 155; Conservative 54; Mismatches 126; Indels 25; Gaps 6;

Qy 38 IVLGIETSCDDTAAAVVDEBTGNVLGEALHSQTEVHLKTGIVPPAAQOLHRENIQRIYOE 97  
Db 55 VVLGIETSCDDTAAAVVSDDELISQVSSQADLLAKTGGVAPKAAEASHSVIDQVVOE 114  
Qy 98 ALSAGVSPDLSAATATTKPGLALSGLSFLQVLGQOLKKPFIPIHMEHALTIRL 157  
Db 115 ALDRAVLEKDLTAVALVITIGPGLSLCLRVGQAKAKIAGPRLPIIGHMEHALVRL 174  
Qy 158 TNK-VEPPFLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLKVARRLSLIGHPEC 216  
Db 175 IEKDLQPFMALISGGHNLVLARDLGQYIQLGTTIDDAIGEAYDKAKVLGLDLR--- 231  
Qy 217 STMSGKALIELHAKQGRFHDIKPPLHAKKCDSPFTGLQ----HYTDKIMKKEKEE 271  
Db 232 --RSGGPAIEKLEAMGNASVYKFSIPMKQHKDCNFSYAGLKTQVRLAESKKIDAKIPIIS 289  
Qy 272 GIEKGQILSSAADIAATVQHTMACHLVYKTRRAILFCQKORDLLPONNAVLVASGVASNF 331  
Db 290 SASNGDRL-SPADIAASFORAVALHLEKCEKRAIDMALMESISFH---LVVSGVASNQ 345  
Qy 332 YIRARLEILTNAQCTLLCPPEPLCTDNGIMIANNGIERLRAAGILHDIGIRYEPKCP 391  
Db 346 YVRRARLDMVVKNGQLQVCPPEPLCTDNGVMIAMGIEHFRNG-----RYDPPPP 395

## RESULT 13

US-10-067-443-3  
; Sequence 3, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443

; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266, 518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282, 814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 463  
; TYPE: PRP  
; ORGANISM: Arabidopsis thaliana  
US-10-067-443-3

Query Match 29.8%; Score 634; DB 14; Length 463;  
Best Local Similarity 37.9%; Pred. No. 4.3e-53;  
Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;

Qy 38 IVLGIETSCDDTAAAVVDEBTGNVLGEALHSQTEVHLKTGIVPPAAQOLHRENIQRIYOE 97  
Db 85 VVLGIETSCDDTAAAVVSPFNHSSC---RAELLVQGVGAPAKQAEASHSVIDKVVOD 141  
Qy 98 ALSAGVSPDLSAATATTKPGLALSGLSFLQVLGQOLKKPFIPIHMEHALTIRL 157  
Db 142 ALDKANLIEKDLASAAVITIGPGLSLCLRVGAKARVAGNFSLLPIGVHMEHALVRL 201  
Qy 158 T-NKVEPPFLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLKVARRLSLIGHPEC 216  
Db 202 VQELSPFPMALLISGGHNLVLARDLGQYIQLGTTIDDAIGEAYDKAKVLGLDLR--- 258  
Qy 217 STMSGKALIELHAKQGRFHDIKPPLHAKKCDSPFTGLQOHTYTDKIMKKEKEEIEKG 276  
Db 259 --RSGGPAIEBELADGASVAFNPMKYKDCNFSYAGLKTQVRLAIEKX----- 308  
Qy 277 QILSSAADIAATVQHTMACHLVYKTRRAILFCQKORDLLPONNAVLVASGVASNFIRRA 336  
Db 309 --IRNRADIAASFQRAVALHLEKCEKRAIDMALE--LEPSIKHWISGVASNKYVRLR 363  
Qy 337 IELTNAQCTLLCPPEPLCTDNGIMIANNGIERLRAAGILHDIGIRYE----- 387  
Db 364 LNNIVENKMLKLVCPPEPLCTDNGVAVWATGIEHFRVQ-----RYDPPPATPEE 413  
Qy 388 -----PKCPLGVDISKVEGEA 403  
Db 414 DYVYDLRPRWPLGEBYAKGRSEA 436

## RESULT 14

US-10-649-273-3  
; Sequence 3, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; CURRENT FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266, 518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282, 814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 463  
; TYPE: PRP  
; ORGANISM: Arabidopsis thaliana  
US-10-649-273-3

Query Match 29.8%; Score 634; DB 15; Length 463;  
Best Local Similarity 37.9%; Pred. No. 4.3e-53;  
Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;

```

QY 38 VLGJETSDDPAAAVVDSTGVVLGBAHSOFVHLKGTGIVPPAQQOLRENIQRIVOE 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 VVLGJETSDDPAAAVVSPFNHSSC---RAELVQVGVAPOKQBEAHSRTDKVQD 144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 98 ALSASGVSPDSIAIATTIKPIALSLGVLSFSLQVLVQOLKKPFPINHEAHATIRL 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 142 ALDKANLITEKOLSAVAVITIGFGLSLCLRVGVKARARVAAGNFSPIYGVNHEAHATIRL 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 158 T-NKKEPFPVLITISGGHCLALVQGVSPFLIGKSLDIAFGMDKVARRLSLIKHEP 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 202 VEOELSPFPFMAILISGHNHLVLAHLKGYTQGLTVDDAIGRAPKTAAMLGIQDMH--- 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 217 STMSSGKKIIEHLAKQGNRFHPDIKPIHLNAKQDPSFGTGOHYTIDKIIMKKEBGEIENG 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 --RSGGPAVEELATSGDAKSAYFKNPMKHKQCNFSYAGLKTQVRLAIEKE----- 308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 277 QILSSAADIATVQHTMACHLVKTRTHRAILFCQKODLLEPONNAVLVYASGVASNFYIRRA 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 --IRNRADIASFQVAVVHLBEKCEBALDWMLE---LEPSIGHWVITSGSVANNKYVRLR 363
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 337 LEILTNALQOCTLLCEPPRLCTDNGIMIANGLIERLPAIGIILHDIGIRYE----- 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 364 LNNIVENKRLKLVCPSPSLCTDNGVAVVAVTGLEHFFVG-----RYDPPPPATEPE 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 388 -----PKCPLGVADISKEVGEA 403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 414 DYVDLRRPWPJGEEVAKGRSEA 436
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 15
US-10-651-722-3
; Sequence 3, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/566,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-651-722-3

```

Query Match	29.8%;	Score 634;	DB 15;	Length 463;
Best Local Similarity	37.9%;	Pred. No. 4.3e-53;		
Matches 145;	Conservative 61;	Mismatches 129;	Indels 48;	Gaps 7

```

QY      98 ALSASGVSPDLSAIATTIKPGLALSLGVLSPSLOLVGOLKKPPIPIHMHMAHALTIRL  157
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      142 ALDKRNLTEKODLSAVAVTITGPCLSLCRGVARKARVACNFSPIYGVHMHMAHALVARL  201
QY      158 T-NKVEFPFVLLISGGHCLLALVGVSDFFLLAGSLDIAPGMDLDRVARSLIIGHPEC  216
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      202 VEOEISPFPMALLISGGHLLVLAHKLCAQYOLGTVDDAIGAPAPKTAKKWGLIDNH---  258
QY      217 STMSGKRIEHLAKQGNRFHPIDKPIYHNAKCDSPFLGQHTVDKIINKKEBEQIEKG  276
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      259 --NSGGPRAVEEIALTEGDKSVKENVPMKYHKNQCSYVALKQVRLAIRK-----  308

```

[illegible]

Search completed: February 16, 2005, 13:28:29  
Job time : 242.85 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 13:02:52 ; Search time 226.8 Seconds  
(without alignments)  
2132.075 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125  
Sequence: 1 MLILTTAGVFPKRSKRVY.....DISKVGSAIKVPQMKMEI 414

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 6959266 seqs, 116806243 residues

Total number of hits satisfying chosen parameters: 6959266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

- Pending Patents AA Main:\*
- 1: /cgn2\_6/prodata/1/paa/PCYUS COMB.pep.\*
  - 2: /cgn2\_6/prodata/1/paa/US06 COMB.pep.\*
  - 3: /cgn2\_6/prodata/1/paa/US07 COMB.pep.\*
  - 4: /cgn2\_6/prodata/1/paa/US08 COMB.pep.\*
  - 5: /cgn2\_6/prodata/1/paa/US081 COMB.pep.\*
  - 6: /cgn2\_6/prodata/1/paa/US082 COMB.pep.\*
  - 7: /cgn2\_6/prodata/1/paa/US083 COMB.pep.\*
  - 8: /cgn2\_6/prodata/1/paa/US084 COMB.pep.\*
  - 9: /cgn2\_6/prodata/1/paa/US085 COMB.pep.\*
  - 10: /cgn2\_6/prodata/1/paa/US086 COMB.pep.\*
  - 11: /cgn2\_6/prodata/1/paa/US087 COMB.pep.\*
  - 12: /cgn2\_6/prodata/1/paa/US088 COMB.pep.\*
  - 13: /cgn2\_6/prodata/1/paa/US089 COMB.pep.\*
  - 14: /cgn2\_6/prodata/1/paa/US090 COMB.pep.\*
  - 15: /cgn2\_6/prodata/1/paa/US091 COMB.pep.\*
  - 16: /cgn2\_6/prodata/1/paa/US092 COMB.pep.\*
  - 17: /cgn2\_6/prodata/1/paa/US093 COMB.pep.\*
  - 18: /cgn2\_6/prodata/1/paa/US094 COMB.pep.\*
  - 19: /cgn2\_6/prodata/1/paa/US095 COMB.pep.\*
  - 20: /cgn2\_6/prodata/1/paa/US096 COMB.pep.\*
  - 21: /cgn2\_6/prodata/1/paa/US097A COMB.pep.\*
  - 22: /cgn2\_6/prodata/1/paa/US097B COMB.pep.\*
  - 23: /cgn2\_6/prodata/1/paa/US098 COMB.pep.\*
  - 24: /cgn2\_6/prodata/1/paa/US099A COMB.pep.\*
  - 25: /cgn2\_6/prodata/1/paa/US099B COMB.pep.\*
  - 26: /cgn2\_6/prodata/1/paa/US100 COMB.pep.\*
  - 27: /cgn2\_6/prodata/1/paa/US101 COMB.pep.\*
  - 28: /cgn2\_6/prodata/1/paa/US102 COMB.pep.\*
  - 29: /cgn2\_6/prodata/1/paa/US103 COMB.pep.\*
  - 30: /cgn2\_6/prodata/1/paa/US104 COMB.pep.\*
  - 31: /cgn2\_6/prodata/1/paa/US105 COMB.pep.\*
  - 32: /cgn2\_6/prodata/1/paa/US106 COMB.pep.\*
  - 33: /cgn2\_6/prodata/1/paa/US107 COMB.pep.\*
  - 34: /cgn2\_6/prodata/1/paa/US108 COMB.pep.\*
  - 35: /cgn2\_6/prodata/1/paa/US109 COMB.pep.\*
  - 36: /cgn2\_6/prodata/1/paa/US110 COMB.pep.\*
  - 37: /cgn2\_6/prodata/1/paa/US60 COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	414	1 PCT-US02-03353-2	Sequence 2, Appl1
2	2125	100.0	414	1 PCT-US02-19360-8	Sequence 8, Appl1
3	2125	100.0	414	30 US-10-480-988-8	Sequence 2, Appl1
4	2125	100.0	414	30 US-10-649-273-2	Sequence 2, Appl1
5	2125	100.0	414	32 US-10-651-722-2	Sequence 2, Appl1
6	2125	100.0	414	32 US-60-266-518-2	Sequence 2, Appl1
7	2125	100.0	414	37 US-60-282-814-2	Sequence 2, Appl1
8	2125	100.0	425	35 US-10-918-754-1372	Sequence 1372, Ap
9	2125	100.0	425	37 US-60-495-114-1372	Sequence 1372, Ap
10	2090.5	98.4	439	1 PCT-US02-03353-19	Sequence 19, Appl
11	2090.5	98.4	439	32 US-10-651-722-19	Sequence 19, Appl
12	2090.5	98.4	439	32 US-10-651-722-19	Sequence 19, Appl
13	2090.5	98.4	439	37 US-60-266-518-19	Sequence 19, Appl
14	2090.5	98.4	439	37 US-60-282-814-19	Sequence 19, Appl
15	2088	98.3	414	26 US-10-012-140-5	Sequence 2039, Ap
16	1845	86.8	364	26 US-10-094-749-2039	Sequence 2039, Ap
17	1395	65.6	298	27 US-10-170-205E-27317	Sequence 27317, A
18	1395	65.6	298	35 US-10-918-754-1371	Sequence 1371, Ap
19	1395	65.6	298	37 US-60-495-114-1371	Sequence 1371, Ap
20	1395	65.6	309	35 US-10-918-754-1374	Sequence 1374, Ap
21	1395	65.6	309	35 US-10-918-754-1375	Sequence 1375, Ap
22	1395	65.6	309	37 US-60-495-114-1374	Sequence 1374, Ap
23	1395	65.6	309	37 US-60-495-114-1375	Sequence 1375, Ap
24	1385	65.2	267	1 PCT-US02-03353-22	Sequence 22, Appl
25	1385	65.2	267	32 US-10-649-273-22	Sequence 22, Appl
26	1385	65.2	267	35 US-10-651-722-22	Sequence 22, Appl
27	1385	65.2	267	37 US-10-918-754-1373	Sequence 1373, Ap
28	1385	65.2	267	37 US-60-266-518-12	Sequence 22, Appl
29	1385	65.2	267	37 US-60-282-814-22	Sequence 22, Appl
30	1385	65.2	267	37 US-60-495-114-1373	Sequence 1373, Ap
31	827	38.9	291	37 US-60-243-468-1011	Sequence 1011, Ap
32	827	38.9	291	37 US-60-243-742-177	Sequence 177, App
33	714.5	33.6	408	37 US-60-173-464-26192	Sequence 26192, A
34	714.5	33.6	409	20 US-09-614-150-34191	Sequence 34191, A
35	714.5	33.6	409	20 US-09-614-150A-34191	Sequence 34191, A
36	714.5	33.6	409	37 US-60-191-637-33775	Sequence 33775, A
37	714.5	33.6	409	37 US-60-191-681-26629	Sequence 26629, A
38	681.5	32.1	445	30 US-10-424-599-209259	Sequence 209259, A
39	659.5	31.0	439	19 US-09-513-996A-21029	Sequence 21029, A
40	659.5	31.0	444	19 US-09-513-996A-21029	Sequence 21029, A
41	656	30.9	461	30 US-10-449-902-48987	Sequence 48987, A
42	634	29.8	463	1 PCT-US02-03353-3	Sequence 3, Appl1
43	634	29.8	463	21 US-09-708-427-16627	Sequence 16627, A
44	634	29.8	463	32 US-10-649-273-3	Sequence 3, Appl1
45	634	29.8	463	32 US-10-651-722-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1  
PCT-US02-03353-2  
Sequence 2, Application PC/TUS0203353  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
FILE REFERENCE: D0073 PCT  
CURRENT APPLICATION NUMBER: PCT/US02/03353  
PRIOR FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266, 518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282, 814  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: Patcraft version 3.0  
SEQ ID NO 2  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Homo sapiens

PCT-US02-03353-2

Query Match 100.0%; Score 2125; DB 1; Length 414;

Best Local Similarity 100.0%; Pred. No. 1.2e-214; Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLITKTAGVFPKPSKRKYEFRLSFNFHFGTLFLHKIVLGIETSCDDTAAAVDEGTNV 60
DB 1 MLITKTAGVFPKPSKRKYEFRLSFNFHFGTLFLHKIVLGIETSCDDTAAAVDEGTNV 60
QY 61 LGEAHSQTEVHLKGTGIVPPAAOQLHRENIQRIYOEALASAGVSPSLSAITTKPGL 120
DB 61 LGEAHSQTEVHLKGTGIVPPAAOQLHRENIQRIYOEALASAGVSPSLSAITTKPGL 120
QY 121 ALSGVGSPSLQVGLQKPFPIPHHMAHALTLRLTNKVEPFLVLLISGHCILLAV 180
DB 121 ALSGVGSPSLQVGLQKPFPIPHHMAHALTLRLTNKVEPFLVLLISGHCILLAV 180
QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKALEHLAKQGNRFHFDIK 240
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKALEHLAKQGNRFHFDIK 240
QY 241 PPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMACHLYR 300
DB 241 PPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMACHLYR 300
QY 301 THRALIFCKORDLPPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
DB 301 THRALIFCKORDLPPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGIGIILHDIGIRYBPCPLGVDISKVEGASIVPOLKMEI 414
DB 361 IMIANGIERLRAGIGIILHDIGIRYBPCPLGVDISKVEGASIVPOLKMEI 414

```

RESULT 2

PCT-US02-19360-8

Sequence 8, Application PC/TUS0219360

GENERAL INFORMATION:

```

; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: KABLE, Amy E.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: HARALTA, April J.A.
; APPLICANT: TRAN, Bao
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: WARREN, Bridget A.
; APPLICANT: ISON, Craig H.
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: NGUYEN, Daniel B.
; APPLICANT: LU, Dzung Anna M.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: YU, Henry
; APPLICANT: FORSYTHE, Ian J.
; APPLICANT: BARROSO, Ines
; APPLICANT: BARKIMAR, Jayalaxmi
; APPLICANT: GRIFPIN, Jennifer A.
; APPLICANT: LI, Joana X.
; APPLICANT: YANG, Junning
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: GIETZEN, Kimberly J.
; APPLICANT: DING, Li
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: YAO, Monique G.
; APPLICANT: WALIA, Narinder K.
; APPLICANT: MASON, Patricia M.
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: LEE, Sally
; APPLICANT: BECHA, Shanya D.
; APPLICANT: LEE, Soo Yeun
; APPLICANT: TRAN, Uyen K.
; APPLICANT: ELLIOTT, Vicki S.

```

```

; APPLICANT: LUO, Wen
; APPLICANT: SPRAGUE, William
; APPLICANT: TANG, Y. Tom
; APPLICANT: LU, Yan
; APPLICANT: ZEBARADIAN, Yeganeh
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-1040 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/19360
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/300,508
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/303,445
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/305,405
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/311,442
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/314,821
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/315,992
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 60/378,205
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7632424CD1
PCT-US02-19360-8

```

Query Match 100.0%; Score 2125; DB 1; Length 414;

Best Local Similarity 100.0%; Pred. No. 1.2e-214; Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLITKTAGVFPKPSKRKYEFRLSFNFHFGTLFLHKIVLGIETSCDDTAAAVDEGTNV 60
DB 1 MLITKTAGVFPKPSKRKYEFRLSFNFHFGTLFLHKIVLGIETSCDDTAAAVDEGTNV 60
QY 61 LGEAHSQTEVHLKGTGIVPPAAOQLHRENIQRIYOEALASAGVSPSLSAITTKPGL 120
DB 61 LGEAHSQTEVHLKGTGIVPPAAOQLHRENIQRIYOEALASAGVSPSLSAITTKPGL 120
QY 121 ALSGVGSPSLQVGLQKPFPIPHHMAHALTLRLTNKVEPFLVLLISGHCILLAV 180
DB 121 ALSGVGSPSLQVGLQKPFPIPHHMAHALTLRLTNKVEPFLVLLISGHCILLAV 180
QY 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKALEHLAKQGNRFHFDIK 240
DB 181 QGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKALEHLAKQGNRFHFDIK 240
QY 241 PPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMACHLYR 300
DB 241 PPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIEKGQILSSADIAATVQHTMACHLYR 300
QY 301 THRALIFCKORDLPPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
DB 301 THRALIFCKORDLPPONNAVIVASGVASNFYIRALLETITNAQTCTLLCPPRLCTDNG 360
QY 361 IMIANGIERLRAGIGIILHDIGIRYBPCPLGVDISKVEGASIVPOLKMEI 414
DB 361 IMIANGIERLRAGIGIILHDIGIRYBPCPLGVDISKVEGASIVPOLKMEI 414

```

RESULT 3

US-10-480-988-8

Sequence 8, Application US/10480988

GENERAL INFORMATION:

```

; APPLICANT: GANDHI, Ameena R.; KABLE, Amy E.;
; APPLICANT: SWARNAKAR, Anita; HARALTA, April J.A.;

```

APPLICANT: TRAN, Bao; DUGAN, Brendan M.;  
 APPLICANT: WARREN, Bridget A.; ISON, Craig H.;  
 APPLICANT: HONCHILL, Cynthia D.; NGUYEN, Danielle B.;  
 APPLICANT: LU, Dzung Aina M.; LEB, Ernestine A.;  
 APPLICANT: YUE, Henry; FORSYTHE, Ian J.;  
 APPLICANT: BARROSO, Ines; RAMKOMAR, Jayalaxmi;  
 APPLICANT: GRIFFIN, Jennifer A.; LI, Joana X.;  
 APPLICANT: YANG, Junning; THANGAVELU, Kavitha;  
 APPLICANT: GIETZEN, Kimberly J.; DING, Li;  
 APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.;  
 APPLICANT: YAO, Monique G.; CHAMLA, Narinder K.;  
 APPLICANT: MASON, Patricia M.; GURURAJAN, Rajagopal;  
 APPLICANT: LEE, Sally; BECHA, Shanya D.;  
 APPLICANT: LEE, Soo Yeun; TRAN, Uyen K.;  
 APPLICANT: ELLIOTT, Vicki S.; LUO, Wen;  
 APPLICANT: SPRAGUE, William W.; TANG, Y. Tom;  
 APPLICANT: LU, Yan; ZEBARADIAN, Yeganeh  
 TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES  
 FILE REFERENCE: PF-1040 USN  
 CURRENT APPLICATION NUMBER: US/10/480,988  
 PRIOR FILING DATE: 2003-12-16  
 PRIOR APPLICATION NUMBER: PCT/US02/19360  
 PRIOR FILING DATE: 2002-06-18  
 PRIOR APPLICATION NUMBER: US 60/300,508  
 PRIOR FILING DATE: 2001-06-22  
 PRIOR APPLICATION NUMBER: US 60/303,445  
 PRIOR FILING DATE: 2001-07-06  
 PRIOR APPLICATION NUMBER: US 60/305,405  
 PRIOR FILING DATE: 2001-07-13  
 PRIOR APPLICATION NUMBER: US 60/311,442  
 PRIOR FILING DATE: 2001-08-09  
 PRIOR APPLICATION NUMBER: US 60/314,821  
 PRIOR FILING DATE: 2001-08-24  
 PRIOR APPLICATION NUMBER: US 60/315,992  
 PRIOR FILING DATE: 2001-08-29  
 PRIOR APPLICATION NUMBER: US 60/378,205  
 PRIOR FILING DATE: 2002-05-03  
 NUMBER OF SEQ ID NOS: 56  
 SOFTWARE: PERL Program  
 SEQ ID NO 8  
 LENGTH: 414  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc feature  
 OTHER INFORMATION: Incyte ID No: 7632424CD1  
 US-10-480-988-8  
 Query Match 100.0%; Score 2125; DB 30; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-214;  
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLITKTAGVFPKSKRKYEFELRSFNFPGTLFLHKIVLGIEISCDPTAAAVVDETGV 60  
 DB 1 MLITKTAGVFPKSKRKYEFELRSFNFPGTLFLHKIVLGIEISCDPTAAAVVDETGV 60  
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQEALSASGVSPSDLSAIAITTKPG 120  
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQEALSASGVSPSDLSAIAITTKPG 120  
 QY 121 ALSGVGLSFSLOLVGQKKPPIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 180  
 DB 121 ALSGVGLSFSLOLVGQKKPPIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 180  
 QY 121 ALSGVGLSFSLOLVGQKKPPIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 180  
 DB 121 ALSGVGLSFSLOLVGQKKPPIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 180  
 QY 181 QGVSDPFLLLGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGGKAI EHLAKQGNRPFIDIK 240  
 DB 181 QGVSDPFLLLGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGGKAI EHLAKQGNRPFIDIK 240  
 QY 241 PPLHAKNCDFSPFTGLQHTVDKIIMKKEKEBGIKQQLISSADIAATVQHTMACHLVKR 300  
 DB 241 PPLHAKNCDFSPFTGLQHTVDKIIMKKEKEBGIKQQLISSADIAATVQHTMACHLVKR 300  
 QY 301 THRAILFCQKORDLLPQNNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360  
 DB 301 THRAILFCQKORDLLPQNNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360

DB 301 THRAILFCQKORDLLPQNNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360  
 QY 361 IMIANNGIERLRAGLILHDIGIRVEPKCPGVDSKEVGASIVPOLKMEI 414  
 DB 361 IMIANNGIERLRAGLILHDIGIRVEPKCPGVDSKEVGASIVPOLKMEI 414  
 RESULT 4  
 US-10-649-273-2  
 Sequence 2, Application US/10649273  
 GENERAL INFORMATION:  
 APPLICANT: Bristol-Myers Squibb Company  
 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
 FILE REFERENCE: D0073 CNT  
 CURRENT APPLICATION NUMBER: US/10/649,273  
 PRIOR FILING DATE: 2003-08-27  
 PRIOR APPLICATION NUMBER: US 60/266,518  
 PRIOR FILING DATE: 2001-02-05  
 PRIOR APPLICATION NUMBER: US 10/067,443  
 PRIOR FILING DATE: 2002-02-05  
 PRIOR APPLICATION NUMBER: US 60/282,814  
 PRIOR FILING DATE: 2001-04-10  
 NUMBER OF SEQ ID NOS: 71  
 SOFTWARE: PatentIn version 3.2  
 SEQ ID NO 2  
 LENGTH: 414  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-649-273-2  
 Query Match 100.0%; Score 2125; DB 32; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-214;  
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLITKTAGVFPKSKRKYEFELRSFNFPGTLFLHKIVLGIEISCDPTAAAVVDETGV 60  
 DB 1 MLITKTAGVFPKSKRKYEFELRSFNFPGTLFLHKIVLGIEISCDPTAAAVVDETGV 60  
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQEALSASGVSPSDLSAIAITTKPG 120  
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQEALSASGVSPSDLSAIAITTKPG 120  
 QY 121 ALSGVGLSFSLOLVGQKKPPIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 180  
 DB 121 ALSGVGLSFSLOLVGQKKPPIPIHMEAHALITRLTNKVEPPLVLLISGHCILALV 180  
 QY 181 QGVSDPFLLLGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGGKAI EHLAKQGNRPFIDIK 240  
 DB 181 QGVSDPFLLLGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGGKAI EHLAKQGNRPFIDIK 240  
 QY 241 PPLHAKNCDFSPFTGLQHTVDKIIMKKEKEBGIKQQLISSADIAATVQHTMACHLVKR 300  
 DB 241 PPLHAKNCDFSPFTGLQHTVDKIIMKKEKEBGIKQQLISSADIAATVQHTMACHLVKR 300  
 QY 301 THRAILFCQKORDLLPQNNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360  
 DB 301 THRAILFCQKORDLLPQNNNAVIVASGVASNFYIRRALBITLTAQCTLLCPPRLCTDNG 360  
 QY 361 IMIANNGIERLRAGLILHDIGIRVEPKCPGVDSKEVGASIVPOLKMEI 414  
 DB 361 IMIANNGIERLRAGLILHDIGIRVEPKCPGVDSKEVGASIVPOLKMEI 414  
 RESULT 5  
 US-10-651-722-2  
 Sequence 2, Application US/10651722  
 GENERAL INFORMATION:  
 APPLICANT: Bristol-Myers Squibb Company  
 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
 FILE REFERENCE: D0073 DIV  
 CURRENT APPLICATION NUMBER: US/10/651,722  
 PRIOR FILING DATE: 2003-08-29

; PRIOR APPLICATION NUMBER: US 60/266,518  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: US 10/067,443  
 ; PRIOR FILING DATE: 2002-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/282,814  
 ; PRIOR FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 2  
 ; LENGTH: 414  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; -10-651-722-2

Query Match 100.0%; Score 2125; DB 32; Length 414;

Best Local Similarity 100.0%; Pred. No. 1.2e-214;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRVYEFLLRSFNFHFGTLFLHKIVLGIETSCDDPTAAAVDDEGNV 60  
 DB 1 MLITKTAGVFFPKSKRVYEFLLRSFNFHFGTLFLHKIVLGIETSCDDPTAAAVDDEGNV 60  
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQEALASGVSPSDLSAIAATTIKPGL 120  
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQEALASGVSPSDLSAIAATTIKPGL 120  
 QY 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEFPFLVLLISGHCLLALV 180  
 DB 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEFPFLVLLISGHCLLALV 180  
 QY 181 QGVSDFLILGKSLDIPAGMDLKVARRSLIKHPECSITMSGKAIHHLAKQGRFFHDIK 240  
 DB 181 QGVSDFLILGKSLDIPAGMDLKVARRSLIKHPECSITMSGKAIHHLAKQGRFFHDIK 240  
 QY 241 PPLHAKNCDPFSFTGLQHTVDKIIIMKKEKEGIEKQILSSAADIATVQHTMACHLVKR 300  
 DB 241 PPLHAKNCDPFSFTGLQHTVDKIIIMKKEKEGIEKQILSSAADIATVQHTMACHLVKR 300  
 QY 301 THRALLFCQKQDILLPQNNAVLVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360  
 DB 301 THRALLFCQKQDILLPQNNAVLVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360  
 QY 361 IMIANNGERLRAGGILHDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414  
 DB 361 IMIANNGERLRAGGILHDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

# RESULT 6

US-60-266-518-2

; Sequence 2, Application US/60266518  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bristol-Myers Squibb Company  
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
 ; TITLE OF INVENTION: SPINAL CORD, MP-1  
 ; FILE REFERENCE: D0073 PSP  
 ; CURRENT APPLICATION NUMBER: US/60/266,518  
 ; CURRENT FILING DATE: 2001-02-05  
 ; NUMBER OF SEQ ID NOS: 28  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 414  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; -60-266-518-2

Query Match 100.0%; Score 2125; DB 37; Length 414;

Best Local Similarity 100.0%; Pred. No. 1.2e-214;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRVYEFLLRSFNFHFGTLFLHKIVLGIETSCDDPTAAAVDDEGNV 60  
 DB 1 MLITKTAGVFFPKSKRVYEFLLRSFNFHFGTLFLHKIVLGIETSCDDPTAAAVDDEGNV 60

QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQEALASGVSPSDLSAIAATTIKPGL 120  
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQEALASGVSPSDLSAIAATTIKPGL 120  
 QY 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEFPFLVLLISGHCLLALV 180  
 DB 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEFPFLVLLISGHCLLALV 180  
 QY 181 QGVSDFLILGKSLDIPAGMDLKVARRSLIKHPECSITMSGKAIHHLAKQGRFFHDIK 240  
 DB 181 QGVSDFLILGKSLDIPAGMDLKVARRSLIKHPECSITMSGKAIHHLAKQGRFFHDIK 240  
 QY 241 PPLHAKNCDPFSFTGLQHTVDKIIIMKKEKEGIEKQILSSAADIATVQHTMACHLVKR 300  
 DB 241 PPLHAKNCDPFSFTGLQHTVDKIIIMKKEKEGIEKQILSSAADIATVQHTMACHLVKR 300  
 QY 301 THRALLFCQKQDILLPQNNAVLVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360  
 DB 301 THRALLFCQKQDILLPQNNAVLVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360  
 QY 361 IMIANNGERLRAGGILHDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414  
 DB 361 IMIANNGERLRAGGILHDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414

# RESULT 7

US-60-282-814-2

; Sequence 2, Application US/60282814

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN

; TITLE OF INVENTION: SPINAL CORD, MP-1

; FILE REFERENCE: D0073 PSP1

; CURRENT APPLICATION NUMBER: US/60/282,814

; CURRENT FILING DATE: 2001-04-10

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2

; LENGTH: 414

; TYPE: PRT

; ORGANISM: Homo sapiens

; -60-282-814-2

Query Match 100.0%; Score 2125; DB 37; Length 414;

Best Local Similarity 100.0%; Pred. No. 1.2e-214;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLITKTAGVFFPKSKRVYEFLLRSFNFHFGTLFLHKIVLGIETSCDDPTAAAVDDEGNV 60  
 DB 1 MLITKTAGVFFPKSKRVYEFLLRSFNFHFGTLFLHKIVLGIETSCDDPTAAAVDDEGNV 60  
 QY 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQEALASGVSPSDLSAIAATTIKPGL 120  
 DB 61 LGEAHSQTEVHLKTGGIVPPAAQQLHRENIQRIYQEALASGVSPSDLSAIAATTIKPGL 120  
 QY 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEFPFLVLLISGHCLLALV 180  
 DB 121 ALSIGVLSFSLQVGVOLKKEPFIPIHMEAHALTRLTNKKVEFPFLVLLISGHCLLALV 180  
 QY 181 QGVSDFLILGKSLDIPAGMDLKVARRSLIKHPECSITMSGKAIHHLAKQGRFFHDIK 240  
 DB 181 QGVSDFLILGKSLDIPAGMDLKVARRSLIKHPECSITMSGKAIHHLAKQGRFFHDIK 240  
 QY 241 PPLHAKNCDPFSFTGLQHTVDKIIIMKKEKEGIEKQILSSAADIATVQHTMACHLVKR 300  
 DB 241 PPLHAKNCDPFSFTGLQHTVDKIIIMKKEKEGIEKQILSSAADIATVQHTMACHLVKR 300  
 QY 301 THRALLFCQKQDILLPQNNAVLVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360  
 DB 301 THRALLFCQKQDILLPQNNAVLVASGVASNFIYRRALBITLNAQTCTLLCPPRLCTDNG 360  
 QY 361 IMIANNGERLRAGGILHDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414  
 DB 361 IMIANNGERLRAGGILHDIEGIRYEPKCPGVDSKEVGEASIKVPOLKMEI 414



Db 361 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 8

US-10-918-754-1372

Sequence 1372, Application US/10918754

GENERAL INFORMATION:

APPLICANT: CARCILL, Michele

TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND

TITLE OF INVENTION: USES THEREOF

FILE REFERENCE: C0001480

CURRENT APPLICATION NUMBER: US/10/918.754

CURRENT FILING DATE: 2004-08-16

NUMBER OF SEQ ID NOS: 91238

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1372

LENGTH: 425

TYPE: PRT

ORGANISM: Homo sapiens

US-10-918-754-1372

Query Match 100.0%; Score 2125; DB 35; Length 425;

Best Local Similarity 100.0%; Pred. No. 1.3e-214;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFFKPSRRKRYEFLRSFNFHPTGLFLHKIVLGIEISCDPTAAAVDDEGNV 60

Db 12 MLILTKTAGVFFKPSRRKRYEFLRSFNFHPTGLFLHKIVLGIEISCDPTAAAVDDEGNV 71

QY 61 LGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRIYGEALSASGVSPSDLSAATTTIKPGL 120

Db 72 LGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRIYGEALSASGVSPSDLSAATTTIKPGL 131

QY 121 ALSIGVLSFSIQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLAY 180

Db 132 ALSIGVLSFSIQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLAY 191

QY 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQNRFPFDIK 240

Db 192 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQNRFPFDIK 251

QY 241 PPLHAKNCDFSTGLQHTVDKIIMKKEKEGIEKGQIILSSAADIAATVQHTMACHLVKR 300

Db 252 PPLHAKNCDFSTGLQHTVDKIIMKKEKEGIEKGQIILSSAADIAATVQHTMACHLVKR 311

QY 301 THRALIFCKQRDILLPNNNAVLVASGVASNFYIRALIELITNAQTCTLLCPPRLCTDNG 360

Db 312 THRALIFCKQRDILLPNNNAVLVASGVASNFYIRALIELITNAQTCTLLCPPRLCTDNG 371

QY 361 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI 414

Db 372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI 425

RESULT 9

US-60-495-114-1372

Sequence 1372, Application US/60495114

GENERAL INFORMATION:

APPLICANT: CARCILL, Michele

TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND

TITLE OF INVENTION: USES THEREOF

FILE REFERENCE: C0001480

CURRENT APPLICATION NUMBER: US/60/495.114

CURRENT FILING DATE: 2003-08-15

NUMBER OF SEQ ID NOS: 91238

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1372

LENGTH: 425

TYPE: PRT

ORGANISM: Homo sapiens

US-60-495-114-1372

Query Match 100.0%; Score 2125; DB 37; Length 425;

Best Local Similarity 100.0%; Pred. No. 1.3e-214;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFFKPSRRKRYEFLRSFNFHPTGLFLHKIVLGIEISCDPTAAAVDDEGNV 60

Db 12 MLILTKTAGVFFKPSRRKRYEFLRSFNFHPTGLFLHKIVLGIEISCDPTAAAVDDEGNV 71

QY 61 LGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRIYGEALSASGVSPSDLSAATTTIKPGL 120

Db 72 LGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRIYGEALSASGVSPSDLSAATTTIKPGL 131

QY 121 ALSIGVLSFSIQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLAY 180

Db 132 ALSIGVLSFSIQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLAY 191

QY 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQNRFPFDIK 240

Db 192 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQNRFPFDIK 251

QY 241 PPLHAKNCDFSTGLQHTVDKIIMKKEKEGIEKGQIILSSAADIAATVQHTMACHLVKR 300

Db 252 PPLHAKNCDFSTGLQHTVDKIIMKKEKEGIEKGQIILSSAADIAATVQHTMACHLVKR 311

QY 301 THRALIFCKQRDILLPNNNAVLVASGVASNFYIRALIELITNAQTCTLLCPPRLCTDNG 360

Db 312 THRALIFCKQRDILLPNNNAVLVASGVASNFYIRALIELITNAQTCTLLCPPRLCTDNG 371

QY 361 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI 414

Db 372 IMIANGIERLRAGLGIHLDIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI 425

RESULT 10

PCT-US02-03353-19

Sequence 19, Application PC/TUS0203353

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1

FILE REFERENCE: D0073 PCT

CURRENT APPLICATION NUMBER: PCT/US02/03353

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.0

SEQ ID NO 19

LENGTH: 439

TYPE: PRT

ORGANISM: homo sapiens

PCT-US02-03353-19

Query Match 98.4%; Score 2090.5; DB 1; Length 439;

Best Local Similarity 93.8%; Pred. No. 6e-211;

Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLILTKTAGVFFKPSRRKRYEFLRSFNFHPTGLFLHKIVLGIEISCDPTAAAVDDEGNV 60

Db 1 MLILTKTAGVFFKPSRRKRYEFLRSFNFHPTGLFLHKIVLGIEISCDPTAAAVDDEGNV 60

QY 61 LGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRIYGEALSASGVSPSDLSAATTTIKPGL 120

Db 61 LGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRIYGEALSASGVSPSDLSAATTTIKPGL 120

QY 121 ALSIGVLSFSIQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLAY 180

Db 121 ALSIGVLSFSIQVLGQKKPFIPIHMEAHALITRLTNKVEPPLVLLISGHCILLAY 180

QY 181 QGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAI EHLAKQNRFPFDIK 240

```

Db      181 QGVSDPFLLGKSLDIAPGMDLKVARRLSLTKHPECSMTSGKAIENLAKQGNRFPFDIK 240
QY      241 PPLHAKNCDFSPFTGLQHTVDKTIIMKKEKEBGI-----EK 275
Db      241 PPLHAKNCDFSPFTGLQHTVDKTIIMKKEKEBGIPLISVEQINIPGLCLKIAHFCRYEK 300
QY      276 GQILSSADIAIATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 335
Db      301 GQILSSADIAIATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 360
QY      336 ALBITLTAQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLGLIHDIEGIRYBPKPLGYD 395
Db      361 ALBITLTAQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLGLIHDIEGIRYBPKPLGYD 420
QY      396 ISKEVGEASIKVPOLKMEI 414
Db      421 ISKEVGEASIKVPOLKMEI 439

RESULT 11
US-10-649-273-19
; Sequence 19, Application US/10649273
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-19

Query Match      98.4%; Score 2090.5; DB 32; Length 439;
Best Local Similarity 93.8%; Pred. No. 6e-211;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY      1 MLILTKTAGVFFPKSKRKVEFLRSFNFHPGTLFLHKIVLGIETSCDDTPAAVVDETGVN 60
Db      1 MLILTKTAGVFFPKSKRKVEFLRSFNFHPGTLFLHKIVLGIETSCDDTPAAVVDETGVN 60
QY      61 LGEAHSOTEVHLKTGGIVPPAAQOLHRENIQRIYOEALASASGVSPDLAATTKPG 120
Db      61 LGEAHSOTEVHLKTGGIVPPAAQOLHRENIQRIYOEALASASGVSPDLAATTKPG 120
QY      121 ALSIGVGSFSLOLVGOLKKPFIPIHMEAHALTLRLTNKVEFPFLVLLISGHCCLALV 180
Db      121 ALSIGVGSFSLOLVGOLKKPFIPIHMEAHALTLRLTNKVEFPFLVLLISGHCCLALV 180
QY      121 ALSIGVGSFSLOLVGOLKKPFIPIHMEAHALTLRLTNKVEFPFLVLLISGHCCLALV 180
Db      121 ALSIGVGSFSLOLVGOLKKPFIPIHMEAHALTLRLTNKVEFPFLVLLISGHCCLALV 180
QY      181 QGVSDPFLLGKSLDIAPGMDLKVARRLSLTKHPECSMTSGKAIENLAKQGNRFPFDIK 240
Db      181 QGVSDPFLLGKSLDIAPGMDLKVARRLSLTKHPECSMTSGKAIENLAKQGNRFPFDIK 240
QY      241 PPLHAKNCDFSPFTGLQHTVDKTIIMKKEKEBGI-----EK 275
Db      241 PPLHAKNCDFSPFTGLQHTVDKTIIMKKEKEBGIPLISVEQINIPGLCLKIAHFCRYEK 300
QY      276 GQILSSADIAIATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 335
Db      301 GQILSSADIAIATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 360
QY      336 ALBITLTAQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLGLIHDIEGIRYBPKPLGYD 395
Db      361 ALBITLTAQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLGLIHDIEGIRYBPKPLGYD 420

```

```

QY      396 ISKEVGEASIKVPOLKMEI 414
Db      421 ISKEVGEASIKVPOLKMEI 439

RESULT 12
US-10-651-722-19
; Sequence 19, Application US/10651722
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-19

Query Match      98.4%; Score 2090.5; DB 32; Length 439;
Best Local Similarity 93.8%; Pred. No. 6e-211;
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY      1 MLILTKTAGVFFPKSKRKVEFLRSFNFHPGTLFLHKIVLGIETSCDDTPAAVVDETGVN 60
Db      1 MLILTKTAGVFFPKSKRKVEFLRSFNFHPGTLFLHKIVLGIETSCDDTPAAVVDETGVN 60
QY      61 LGEAHSOTEVHLKTGGIVPPAAQOLHRENIQRIYOEALASASGVSPDLAATTKPG 120
Db      61 LGEAHSOTEVHLKTGGIVPPAAQOLHRENIQRIYOEALASASGVSPDLAATTKPG 120
QY      121 ALSIGVGSFSLOLVGOLKKPFIPIHMEAHALTLRLTNKVEFPFLVLLISGHCCLALV 180
Db      121 ALSIGVGSFSLOLVGOLKKPFIPIHMEAHALTLRLTNKVEFPFLVLLISGHCCLALV 180
QY      181 QGVSDPFLLGKSLDIAPGMDLKVARRLSLTKHPECSMTSGKAIENLAKQGNRFPFDIK 240
Db      181 QGVSDPFLLGKSLDIAPGMDLKVARRLSLTKHPECSMTSGKAIENLAKQGNRFPFDIK 240
QY      241 PPLHAKNCDFSPFTGLQHTVDKTIIMKKEKEBGI-----EK 275
Db      241 PPLHAKNCDFSPFTGLQHTVDKTIIMKKEKEBGIPLISVEQINIPGLCLKIAHFCRYEK 300
QY      276 GQILSSADIAIATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 335
Db      301 GQILSSADIAIATVOHTMACHLVKRTHRAILFCQKRDLLPNNNAVVASGVASNFYIR 360
QY      336 ALBITLTAQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLGLIHDIEGIRYBPKPLGYD 395
Db      361 ALBITLTAQCTLLCPPEPLCTDNGIMIAMNGIERLRAGLGLIHDIEGIRYBPKPLGYD 420
QY      396 ISKEVGEASIKVPOLKMEI 414
Db      421 ISKEVGEASIKVPOLKMEI 439

RESULT 13
US-60-266-518-19
; Sequence 19, Application US/60266518
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1

```

FILE REFERENCE: D0073 PSP  
CURRENT APPLICATION NUMBER: US/60/266,518  
CURRENT FILING DATE: 2001-02-05  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 19  
LENGTH: 439  
TYPE: PRT  
ORGANISM: homo sapiens  
US-60-266-518-19

Query Match 98.4%; Score 2090.5; DB 37; Length 439;  
Best Local Similarity 93.8%; Pred. No. 6e-211;  
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLILTKTAGVFPKSRKRYEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60  
DB 1 MLILTKTAGVFPKSRKRYEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60  
QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIVOEALSASGVSPDLSAIAATTIKPGL 120  
DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIVOEALSASGVSPDLSAIAATTIKPGL 120  
QY 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALTRLTNKVEPPLVLLISGHCCLALV 180  
DB 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALTRLTNKVEPPLVLLISGHCCLALV 180  
QY 181 QGVSDFLILGKSLDIPAGMDLKVARRLSLIKPEECSTWSGGAIEHLAQGRFHFDIK 240  
DB 181 QGVSDFLILGKSLDIPAGMDLKVARRLSLIKPEECSTWSGGAIEHLAQGRFHFDIK 240  
QY 241 PPLHAKNCDPFTGLQHTVDKTIIMKKEBEGIFLSKVEQINIPGLCLKIAHFCRYEK 275  
DB 241 PPLHAKNCDPFTGLQHTVDKTIIMKKEBEGIFLSKVEQINIPGLCLKIAHFCRYEK 300  
QY 276 GQILSSAADIAATVOHTMACHLVKTRTRAILFCQKRDLLPONNAVIVASGVASNFYIR 335  
DB 301 GQILSSAADIAATVOHTMACHLVKTRTRAILFCQKRDLLPONNAVIVASGVASNFYIR 360  
QY 336 ALEILTNATQCTLLCPPLCTDNGIMIAMNGIERLPAIGLILHDIIEGIRYBKCPLGVD 395  
DB 361 ALEILTNATQCTLLCPPLCTDNGIMIAMNGIERLPAIGLILHDIIEGIRYBKCPLGVD 420  
QY 396 ISKEVGEASIKVPOLKMEI 414  
DB 421 ISKEVGEASIKVPOLKMEI 439

RESULT 14  
US-60-282-814-19

Sequence 19, Application US/60282814  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
TITLE OF INVENTION: SPINAL CORD, MP-1  
FILE REFERENCE: D0073 PSP  
CURRENT APPLICATION NUMBER: US/60/282,814  
CURRENT FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 19  
LENGTH: 439  
TYPE: PRT  
ORGANISM: homo sapiens  
US-60-282-814-19

Query Match 98.4%; Score 2090.5; DB 37; Length 439;  
Best Local Similarity 93.8%; Pred. No. 6e-211;  
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLILTKTAGVFPKSRKRYEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60  
DB 1 MLILTKTAGVFPKSRKRYEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60

QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIVOEALSASGVSPDLSAIAATTIKPGL 120  
DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIVOEALSASGVSPDLSAIAATTIKPGL 120  
QY 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALTRLTNKVEPPLVLLISGHCCLALV 180  
DB 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALTRLTNKVEPPLVLLISGHCCLALV 180  
QY 181 QGVSDFLILGKSLDIPAGMDLKVARRLSLIKPEECSTWSGGAIEHLAQGRFHFDIK 240  
DB 181 QGVSDFLILGKSLDIPAGMDLKVARRLSLIKPEECSTWSGGAIEHLAQGRFHFDIK 240  
QY 241 PPLHAKNCDPFTGLQHTVDKTIIMKKEBEGIFLSKVEQINIPGLCLKIAHFCRYEK 275  
DB 241 PPLHAKNCDPFTGLQHTVDKTIIMKKEBEGIFLSKVEQINIPGLCLKIAHFCRYEK 300  
QY 276 GQILSSAADIAATVOHTMACHLVKTRTRAILFCQKRDLLPONNAVIVASGVASNFYIR 335  
DB 301 GQILSSAADIAATVOHTMACHLVKTRTRAILFCQKRDLLPONNAVIVASGVASNFYIR 360  
QY 336 ALEILTNATQCTLLCPPLCTDNGIMIAMNGIERLPAIGLILHDIIEGIRYBKCPLGVD 395  
DB 361 ALEILTNATQCTLLCPPLCTDNGIMIAMNGIERLPAIGLILHDIIEGIRYBKCPLGVD 420  
QY 396 ISKEVGEASIKVPOLKMEI 414  
DB 421 ISKEVGEASIKVPOLKMEI 439

RESULT 15  
US-10-012-140-5

Sequence 5, Application US/10012140  
GENERAL INFORMATION:  
APPLICANT: Leiby, Kevin R.  
APPLICANT: Kapeller-Libermann, Rosana  
TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
FILE REFERENCE: 381552004900  
CURRENT APPLICATION NUMBER: US/10/012,140  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: 60/246,768  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,772  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,185  
PRIOR FILING DATE: 2000-11-15  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-012-140-5

Query Match 98.3%; Score 2088; DB 26; Length 414;  
Best Local Similarity 98.3%; Pred. No. 1e-210;  
Matches 407; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPKSRKRYEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60  
DB 1 MLILTKTAGVFPKSRKRYEFLRSFNFHPTGLFLHKIVLGIETSCDDTAAAVDDETVN 60  
QY 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIVOEALSASGVSPDLSAIAATTIKPGL 120  
DB 61 LGEAHSQTEVHLKTGIVPPAAQOLHRENIQRIVOEALSASGVSPDLSAIAATTIKPGL 120  
QY 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALTRLTNKVEPPLVLLISGHCCLALV 180  
DB 121 ALSGVLSFSLQVLGQKKPFIPIHMEAHALTRLTNKVEPPLVLLISGHCCLALV 180

Qy	181	QGVSDPILLGKSLDIAPGMDLVARRLSLKHPECSTMSGGKAIENHAKQENRPHFDIK	240
Db	181	QGVSDPILLGKSLDIAPGMDLVARRLSLKHPECSTMSGGKAIENHAKQENRPHFDIK	240
Qy	241	PPLHAAKCDPSFTGLQHTDXTIMKKEKEGIEKGQILSSNADIATVQHTMACHLYR	300
Db	241	PPLHAAKCDPSFTGLQHTDXTIMKKEKEGIEKGQILSSNADIATVQHTMACHLYR	300
Qy	301	THRALFCCKORDLPONNAVIVASGVASNFYIRALFILTMANQCTLLCPPRLCTDNG	360
Db	301	THRALFCCKORDLPONNAVIVASGVASNFYIRALFILTMANQCTLLCPPRLCTDNG	360
Qy	361	IMIAMNGIERLRAGIGILHDIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI	414
Db	361	IMIAMNGIERLRAGIGILHDIEGIRYBPCPLGVDISKEVGEASIKVPOLKMEI	414

Search completed: February 16, 2005, 13:18:55  
 Job time : 229.8 secs







```
Db      64 BALBAGTTEBDVTAVAATVYGVGLVGLVGLSAKAPAMAHGLPLIPVNHMAHGLMAAQ 123
QY      157 LTNKVEPPLVLLISGGCHLALVGVSDPFLLGSLDAPGMDLKVARRSLIKHPEC 216
Db      124 SVEPLEPFLALVSGHTELTVYVSEAGDYKIVGETRDAVGEADYKVRVWGL----- 177
QY      217 STMSGKAIEMHAKOGNRFHFDIKPPLHAKNCDSPFTGLOHVTDKIIMKKEEGIEKG 276
Db      178 -TYAGREIDELAHQODI-YDPPRAMIKEDNLEFSFGSKAFINLHNAA-----QKG 230
QY      277 QILSSADIAATVOHTMACHLVKERTHRAILFCQKRDLLPQNNAVLVASGVASNFYIRRA 336
Db      231 ESLST-EDLCASFOAVMDIIMAKTKKAL-----BEYVVK--TLVAGVAAANKGLRER 281
QY      337 LEILTNAQCTLLCPPLCTDNGIMIA-----WN 366
Db      282 L--AAEITDVKVIIIPRLCGDNAGMIAYAVSEWN 315
```

## RESULT 7

```
US-11-027-879-4857
; Sequence 4857, Application US/11027879
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qidong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweart, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: 3687.1000-021
; CURRENT APPLICATION NUMBER: US/11/027,879
; PRIOR FILING DATE: 2004-12-30
; PRIOR APPLICATION NUMBER: US 10/640,833
; PRIOR FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4857
; LENGTH: 336
; TYPE: PRF
; ORGANISM: Streptococcus pneumoniae
US-11-027-879-4857
```

```
Query Match      21.3%; Score 453; DB 7; Length 336;
Best Local Similarity 34.2%; Pred. No. 1.2e-33;
Matches 115; Conservative 57; Mismatches 134; Indels 30; Gaps 8;
```

```
QY      37 KIVLGIEISCDPTAAAVDGTGNVGEAHSQTEVHLKGTGIVPPAAQOLHRENIQRIYQ 96
Db      4 RYILAFETSCDRTSAVAVLKNDDELSNVASQIESHKRGVPEVASHHHEVITACIE 63
QY      97 EALSASGVSPDLASAIATTIKPGALSLGVGLSFSLOLVGQKKPFIPIHMEAHALTR 156
Db      64 BALBAGTTEBDVTAVAATVYGVGLVGLVGLSAKAPAMAHGLPLIPVNHMAHGLMAAQ 123
QY      157 LTNKVEPPLVLLISGGCHLALVGVSDPFLLGSLDAPGMDLKVARRSLIKHPEC 216
Db      124 SVEPLEPFLALVSGHTELTVYVSEAGDYKIVGETRDAVGEADYKVRVWGL----- 177
QY      217 STMSGKAIEMHAKOGNRFHFDIKPPLHAKNCDSPFTGLOHVTDKIIMKKEEGIEKG 276
Db      178 -TYAGREIDELAHQODI-YDPPRAMIKEDNLEFSFGSKAFINLHNAA-----QKG 230
QY      277 QILSSADIAATVOHTMACHLVKERTHRAILFCQKRDLLPQNNAVLVASGVASNFYIRRA 336
```

```
Db      231 ESLST-EDLCASFOAVMDIIMAKTKKAL-----BEYVVK--TLVAGVAAANKGLRER 281
QY      337 LEILTNAQCTLLCPPLCTDNGIMIA-----WN 366
Db      282 L--AAEITDVKVIIIPRLCGDNAGMIAYAVSEWN 315
```

## RESULT 8

```
US-11-028-149-4857
; Sequence 4857, Application US/11028149
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; APPLICANT: Zeng, Qidong
; APPLICANT: Opperman, Timothy
; APPLICANT: Houseweart, Chad Eric
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: 3687.1000-017
; CURRENT APPLICATION NUMBER: US/11/028,149
; PRIOR FILING DATE: 2004-12-30
; PRIOR APPLICATION NUMBER: US 10/640,833
; PRIOR FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: US 09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4857
; LENGTH: 336
; TYPE: PRF
; ORGANISM: Streptococcus pneumoniae
US-11-028-149-4857
```

```
Query Match      21.3%; Score 453; DB 7; Length 336;
Best Local Similarity 34.2%; Pred. No. 1.2e-33;
Matches 115; Conservative 57; Mismatches 134; Indels 30; Gaps 8;
```

```
QY      37 KIVLGIEISCDPTAAAVDGTGNVGEAHSQTEVHLKGTGIVPPAAQOLHRENIQRIYQ 96
Db      4 RYILAFETSCDRTSAVAVLKNDDELSNVASQIESHKRGVPEVASHHHEVITACIE 63
QY      97 EALSASGVSPDLASAIATTIKPGALSLGVGLSFSLOLVGQKKPFIPIHMEAHALTR 156
Db      64 BALBAGTTEBDVTAVAATVYGVGLVGLVGLSAKAPAMAHGLPLIPVNHMAHGLMAAQ 123
QY      157 LTNKVEPPLVLLISGGCHLALVGVSDPFLLGSLDAPGMDLKVARRSLIKHPEC 216
Db      124 SVEPLEPFLALVSGHTELTVYVSEAGDYKIVGETRDAVGEADYKVRVWGL----- 177
QY      217 STMSGKAIEMHAKOGNRFHFDIKPPLHAKNCDSPFTGLOHVTDKIIMKKEEGIEKG 276
Db      178 -TYAGREIDELAHQODI-YDPPRAMIKEDNLEFSFGSKAFINLHNAA-----QKG 230
QY      277 QILSSADIAATVOHTMACHLVKERTHRAILFCQKRDLLPQNNAVLVASGVASNFYIRRA 336
Db      231 ESLST-EDLCASFOAVMDIIMAKTKKAL-----BEYVVK--TLVAGVAAANKGLRER 281
QY      337 LEILTNAQCTLLCPPLCTDNGIMIA-----WN 366
Db      282 L--AAEITDVKVIIIPRLCGDNAGMIAYAVSEWN 315
```

## RESULT 9

```
US-11-027-802-4857
; Sequence 4857, Application US/11027802
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
```







## RESULT 14

US-11-027-844-4857

; Sequence 4857, Application US/11027844

; GENERAL INFORMATION:

; APPLICANT: Doucette-Stamm, Lynn

; APPLICANT: Bush, David

; APPLICANT: Zeng, Qilandong

; APPLICANT: Opperman, Timothy

; APPLICANT: Housewart, Chad Eric

; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus

; FILE REFERENCE: 3687.1000-012

; CURRENT APPLICATION NUMBER: US/11/027,844

; PRIOR FILING DATE: 2004-12-30

; PRIOR APPLICATION NUMBER: US 10/640,833

; PRIOR FILING DATE: 2003-08-14

; PRIOR APPLICATION NUMBER: US 09/583,110

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/107,433

; PRIOR FILING DATE: 1998-06-30

; PRIOR APPLICATION NUMBER: US 60/085,131

; PRIOR FILING DATE: 1998-05-12

; PRIOR APPLICATION NUMBER: US 60/051,553

; PRIOR FILING DATE: 1997-07-02

; NUMBER OF SEQ ID NOS: 5322

; SEQ ID NO 4857

; LENGTH: 336

; TYPE: PR

; ORGANISM: Streptococcus pneumoniae

US-11-027-844-4857

Query Match 21.3%; Score 453; DB 7; Length 336;

Best Local Similarity 34.2%; Pred. No. 1.2e-33;

Matches 115; Conservative 57; Mismatches 134; Indels 30; Gaps 8;

QY 37 KIVIGTSCDDTAADVDEGNVIGBAIHSOTVEHLKTGIVPPAOLHRENIQRIQV 96

DB 4 RYLAFTSCDETSVAVLKNDDELSNVIASQIESHKRFQGVPEVASRHHEVITACIE 63

QY 97 EALSASGVSPDLSAIAITIKPGLALSLGVLSFSIQVLVQOLKRPPIPIHMEAHALTIR 156

DB 64 EALAEAGITEEDVTAAVATYVPGVGLVGLSAKAFAMAHGRLPIVNHMAGHILMAAQ 123

QY 157 LTNKVEPPVLLVLSGHCGLALVGVSDPFLILKSLDIAPGMDLKVARRLSLKHPEC 216

DB 124 SVEPLFEPLLALVSGHTELIVYSBAGDYKIVGETRDAVGEAYDKGVWGL----- 177

QY 217 STMSGKAIIEHLAKOGRHFHDIKPPLHAKOCDPSFTGLQHYTDKIIMKEKEGIEK 276

DB 178 -TYPAGEIDELAHQGDYI-YDPPRAMIKEDNLEFSFGLSAPINLHNAE-----QKG 230

QY 277 QILSSADIATVQHTMACHLVKTRTHAILFCQORDLLPQNNAVLVASGVASNFYIRRA 336

DB 231 ESLST-EDLCASFQAAVMDILMAKTKAL-----BEYVVK--TLVAVAGVANKGLRER 281

QY 337 LEILTNATQCTLLCPPRLCTDNGIMIA-----WN 366

DB 282 L--AAETDVKVIIPPLRLGDNAGMIAVAVSEWN 315

## RESULT 15

US-11-028-050-4857

; Sequence 4857, Application US/11028050

; GENERAL INFORMATION:

; APPLICANT: Doucette-Stamm, Lynn

; APPLICANT: Bush, David

; APPLICANT: Zeng, Qilandong

; APPLICANT: Opperman, Timothy

; APPLICANT: Housewart, Chad Eric

; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus

; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics

; FILE REFERENCE: 3687.1000-023

; CURRENT APPLICATION NUMBER: US/11/028,050

; CURRENT FILING DATE: 2004-12-30

; PRIOR APPLICATION NUMBER: US 10/640,833

; PRIOR FILING DATE: 2003-08-14

; PRIOR APPLICATION NUMBER: US 09/583,110

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/107,433

; PRIOR FILING DATE: 1998-06-30

; PRIOR APPLICATION NUMBER: US 60/085,131

; PRIOR FILING DATE: 1998-05-12

; PRIOR APPLICATION NUMBER: US 60/051,553

; PRIOR FILING DATE: 1997-07-02

; NUMBER OF SEQ ID NOS: 5322

; SEQ ID NO 4857

; LENGTH: 336

; TYPE: PR

; ORGANISM: Streptococcus pneumoniae

US-11-028-050-4857

Query Match 21.3%; Score 453; DB 7; Length 336;

Best Local Similarity 34.2%; Pred. No. 1.2e-33;

Matches 115; Conservative 57; Mismatches 134; Indels 30; Gaps 8;

QY 37 KIVIGTSCDDTAADVDEGNVIGBAIHSOTVEHLKTGIVPPAOLHRENIQRIQV 96

DB 4 RYLAFTSCDETSVAVLKNDDELSNVIASQIESHKRFQGVPEVASRHHEVITACIE 63

QY 97 EALSASGVSPDLSAIAITIKPGLALSLGVLSFSIQVLVQOLKRPPIPIHMEAHALTIR 156

DB 64 EALAEAGITEEDVTAAVATYVPGVGLVGLSAKAFAMAHGRLPIVNHMAGHILMAAQ 123

QY 157 LTNKVEPPVLLVLSGHCGLALVGVSDPFLILKSLDIAPGMDLKVARRLSLKHPEC 216

DB 124 SVEPLFEPLLALVSGHTELIVYSBAGDYKIVGETRDAVGEAYDKGVWGL----- 177

QY 217 STMSGKAIIEHLAKOGRHFHDIKPPLHAKOCDPSFTGLQHYTDKIIMKEKEGIEK 276

DB 178 -TYPAGEIDELAHQGDYI-YDPPRAMIKEDNLEFSFGLSAPINLHNAE-----QKG 230

QY 277 QILSSADIATVQHTMACHLVKTRTHAILFCQORDLLPQNNAVLVASGVASNFYIRRA 336

DB 231 ESLST-EDLCASFQAAVMDILMAKTKAL-----BEYVVK--TLVAVAGVANKGLRER 281

QY 337 LEILTNATQCTLLCPPRLCTDNGIMIA-----WN 366

DB 282 L--AAETDVKVIIPPLRLGDNAGMIAVAVSEWN 315

Search completed: February 16, 2005, 13:19:29  
Job time : 15.15 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:10:32 ; Search time 6588 Seconds  
(without alignments)  
3045.000 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 2125  
Sequence: 1 MLILTKAGVFPKSKRKVY.....DISKEVGASIKVPLKMEI 414

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4708233 segs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10649273/runat\_14022005\_114702\_16399/app\_query.fasca\_1.1429  
-DB=GenBank -QFMT=faetap -SUFFIX=ige -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273 @CGC 1.1 8655 @runat\_14022005\_114702\_16399 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenBank.\*  
1: gb\_ba.\*  
2: gb\_hg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	1908	9	BC011904 Homo sapi
2	2125	100.0	2197	6	AR428803 Sequence
3	2090.5	98.4	1387	6	AR428808 Sequence
4	2090.5	98.4	1387	9	HS4295148 Homo sapi

5	2088	98.3	1245	6	AX664697 Sequence
6	2088	98.3	1820	6	AX664695 Sequence
7	1944	91.5	2208	6	AX713716 Sequence
8	1944	91.5	2208	9	AK055441 Homo sapi
9	1835	86.4	1844	10	BC058172 Homo sapi
10	1747	82.2	1416	6	AR541929 Sequence
11	1725	81.2	1546	10	BC078974 Rattus no
12	1385	65.2	1556	6	AR428809 Sequence
13	1362	64.1	1522	5	BX934991 Gallus ga
14	1273	59.9	1017	10	BC038910 Mus muscu
15	1208	56.8	1558	5	BX930963 Gallus ga
16	1193.5	56.2	8415	6	AR428807 Sequence
17	1186.5	55.8	14364	9	AR428807 Sequence
18	995.5	46.8	249601	2	AC114153 Rattus no
19	995.5	46.8	308652	2	AC121478 Rattus no
20	970	45.6	256751	10	AC122925 Mus muscu
21	950.5	44.7	1109	10	BC051211 Mus muscu
22	938	44.1	860	5	BX930694 Gallus ga
23	827	38.9	875	6	CQ721898 Sequence
24	751.5	35.4	121251	5	AL591593 zebrafish
25	714.5	33.6	1576	3	AY051882 Drosophila
26	714.5	33.6	1601	6	CQ606432 Sequence
27	710	33.4	1474	3	AK113378 Clona int
28	709	33.4	117322	5	AL672217 zebrafish
29	700.5	33.0	1385	6	BD157102 Primer fo
30	700.5	33.0	1385	6	AX878239 Sequence
31	700.5	33.0	1385	9	AK027836 Homo sapi
32	687.5	32.4	3656	6	CQ606431 Sequence
33	687.5	32.4	14679	2	AC018262 Drosophila
34	687.5	32.4	180263	3	AC010671 Drosophila
35	687.5	32.4	207432	3	AE003513 Drosophila
36	668.5	31.5	1443	3	AY024338 Arabidops
37	668.5	31.5	1474	8	AY117283 Arabidops
38	668.5	31.5	1567	8	AY063864 Arabidops
39	662.5	31.2	1557	8	AY084577 Arabidops
40	656	30.9	1672	8	AK070912 Oryza sat
41	629	29.6	571	6	BD154975 Oryza sat
42	629	29.6	571	6	AX874913 Sequence
43	618	29.1	110000	2	AP006501_08 Continuation (9 of
44	577.5	27.2	260271	1	AB017258 Wolbachia
45	568.5	26.8	333800	1	SM591792 Sinorhizo

## ALIGNMENTS

RESULT 1	BC011904	1908 bp	mRNA	linear	PRI 23-DEC-2003
LOCUS	BC011904				
DERIVATION	Homo sapiens O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA clone MGC:20293 IMAGE:4121450), complete cds.				
ACCESSION	BC011904				
VERSION	BC011904.2	GI:40225818			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 1908) Strausberg,R.L., Feingold,R.A., Grouse,L.H., Derge,J.G., Klapper,R.D., Collins,P.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.P., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,D., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.B., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullan,S.J., Bosak,S.A., McEwan,P.A., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hultk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,B., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shenvikeno,Y., Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,				

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REMARK  
COMMENT

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Scherf, A., Schein, J.B., Jones, S.J., and Marra, M.A.,  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
2 (bases 1 to 1908)  
Strasberg, R.  
Direct Submission  
Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Dec 19, 2003 this sequence version replaced gi:15080281.  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Issue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@hgrl.nih.gov](mailto:nisc_mgc@hgrl.nih.gov)  
Ahnter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masfello, C., Maskeri, B., Mastrian, S.D., McCluskey, J.C., McDowell, J., Pearson, R., Stantirpop, S., Thomas, P.J., Touchman, J.W., Taurgeon, C., Vogt, J.L., Walker, M.A., Weethey, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNLT at: <http://image.llnl.gov>  
Series: IRAL Plate: 28 Row: 1 Column: 22.  
Location/Qualifiers  
1. 1908  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:20293 IMAGE:4121450"  
/tissue\_type="Muscle; rhabdomyosarcoma"  
/clone\_id="NIH\_MGC\_17"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
1. 1908  
/gene="OSGEP1"  
/db\_xref="LocusID:64172"  
104. 1348  
/gene="OSGEP1"  
/codon\_start=1  
/product="OSGEP1 protein"  
/protein\_id="AAH1904.1"  
/db\_xref="GI:15080282"  
/translation="MLILKTAGVFPKPKRKYEFLLSPNFPTLFLHKIVLGIETSCDPTAAVDETNVGEAHSOFEVLTGKGIYPPAOLHREINIRIYVEALASGSPSDLSAIVTTIKGLAISLVGLSPSLDNLGOLKRPPIPIHMEHNAALTIIRLTKVVEPFLVLLISGCHLALVGVSDFLKLGSLDIPADMDKARSLIKHPICSTMSGGKAEHLAKQGNRPFPDIKPLHAKNDPSTFTGQHYTDKIMKEKEEGIEKGQISSAADIAVQHTWACHLVKRTRAILTCCKGDLDPONNAVIVASGGVANSFYIRFALEILTNATQCTLLCPPLCTDNGIMIWNGLIRLRAAGAILHDIIGIRYEPKPLGVDSIKVEGASIVPOLKMEI"  
215. 1111  
/gene="OSGEP1"  
/note="Peptidase\_M22; Region: Glycoprotease family"  
/db\_xref="CDD:pfam00814"

ORIGIN

Alignment Scores:

Pred. No.:	6,176-171	Length:	1908
Score:	2125.00	Matches:	414
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0
US-10-649-273-2 (1-414) x BC011904 (1-1908)			
QY	1 MetLeuIleLeuThrThyThrAlaGlyValRhePheIysProSerIysArgIysValTyr	20	
DB	104 AGCTAAATCTTTCACCTAAGCTCAGAGATTTTAAACCAATCAAAAGAAAGTTTAT	163	
QY	21 GluPheLeuArgSerPheAspPheHisProGlyIYThrLeuPheLeuHisIysIleValLeu	40	
DB	164 GAATTTTAAAGATTTTAAATTTTCACTCGAAGACATATTTCTTCAATAAATAGATTG	223	
QY	41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal	60	
DB	224 GGAATTTAAAGATTTTAAATTTTCACTCGAAGACATATTTCTTCAATAAATAGATTG	283	
QY	61 LeuGlyGluAlaIleHisSerGlnThrGluValHisIleuLeuThrGlyGlyIleValPro	80	
DB	284 TTGGAGAGAAACAATCAATCCAACTGAAGTTCAATTTAAACAGGTGGATTTGTCCT	343	
QY	81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer	100	
DB	344 CAGCAGCTCAACAGCTTACAGAGAAATATTTCAAGAAATGACAGAGAGCTTTTCT	403	
QY	101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleuSerProGlyLeu	120	
DB	404 GCCAGTGAAGTCTCTCCAGTACCTTCAGCATTTGCAACATACATTAACAGAGACTT	463	
QY	121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIlyIys	140	
DB	464 GCTTTAAGCTGAGAGTGGCTTATCATTTAGCTTACACTGATGAGACAGTTAAAG	523	
QY	141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys	160	
DB	524 CCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT	563	
QY	161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal	180	
DB	584 GNGAATTTCTTTTATGTTTCTTTGATTTCTGAGAGTCACTGCTGTTGGATTAAGTT	643	
QY	181 GlnGlyValSerAspPheLeuLeuLeuGlyIysSerLeuAspIleAlaProGlyAspMet	200	
DB	644 CAAGAGTTTCAGATTTTCTGCTTCTTGGAAGCTTTTGACATAGCACAGGTGACATG	703	
QY	201 LeuAspIysValAlaArgArgLeuSerLeuIleIysHisProGluCysSerThrMetSer	220	
DB	704 CTTGACAAAGTGGCAAGAGACTTTCTTTAAATTAACATCCAGAGCTCCACATGAGT	763	
QY	221 GlyGlyIysAlaIleGluHisIleuAlaIysGlnGlyAsnArgPheHisAspAspIleIys	240	
DB	764 GGTGGGAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTCATTTGACATCAAA	823	
QY	241 ProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr	260	
DB	824 CCTCCCTTCATCAGCTAAATATGATTTTCTTTTACTGACATTCACACGCTTACT	883	
QY	261 AspIysIleIleMetIysIysGluIysGluGlyIleGluIysGlyGlnIleLeuSer	280	
DB	884 GATTAATTAATATATGAAAAAGAAAGAAAGATTTGAGAAAGGCAAAATCCGTCT	943	
QY	281 SerAlaAlaAspIleAlaAlaThrValGlnHisSerMetAlaCysHisIleuValIysArg	300	
DB	944 TCAGCAGCAGACATTTGCTCCACAGACGACACACATGCAATGATCTTGTGAAGA	1003	
QY	301 ThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnAsnAlaVal	320	
DB	1004 ACAATCGGCTATTTCTGTTTGTAAAGACAGAGACTTCTTACCTCAAAATATGACATG	1063	

QY 321 LeuValAlaSerGlyValAlaSerAspMetLeuValAlaSerGlyValLeu 340  
DB 1064 CTGGTCATCTGGTGGTGGCAAGTAATTTATATCCGACAGGCTCTGGAAATTTTA 1123  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProAlaGluCysThrAspAsnGly 360  
DB 1124 ACAACGCAACACAGTGCATCTTGTGTGTCTCTCCACAGATATGACATGATATATGGC 1183  
QY 361 IleMetIleAlaTPanGlyIleGluThrLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
DB 1184 ATTATGATTCATGAGTAATGATGATTAAGAAAGCTACGCTGCTGGCAATTTTACATGAC 1243  
QY 381 IleGluGlyIleArgGlyArgProGlyCysProLeuGlyValAspIleSerGlyVal 400  
DB 1244 ATGAAAGGCATCCGCTATGAAACCAAAATGTCCTCTTGGAGTAAATATCAAAAGAACTT 1303  
QY 401 GlyGluAlaSerIleValProGluLeuLeuMetGluIle 414  
DB 1304 GGAGAAAGCTTCATMAAAGTACCAAAATTAATAATGAGATA 1345

RESULT 2  
AR428803 2197 bp DNA linear PAT 18-DEC-2003  
LOCUS AR428803 Sequence 1 from patent US 6642041.  
DEFINITION AR428803  
ACCESSION AR428803  
VERSION AR428803.1 GI:40186589  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2197)  
AUTHORS Chen, J., Feder, J. N., Nelson, T. C., Krystek, S. R. and Duclos, F.  
TITILE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;  
FEATURES  
source 1. .2197  
location/Qualifiers  
1. .2197  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Alignment Scores:  
Pred. No.: 7.37e-171 Length: 2197  
Score: 2125.00 Matches: 414  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-10-649-273-2 (1-414) x AR428803 (1-2197)

QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGluLeuValGlyLeuLeuGly 140  
DB 591 GCTTTAAGCTCGGAGTGGGCTTATCATTTAGCTTACACTGGTAGACAGTTAAAAAG 650  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnGly 160  
DB 651 CCATTCATCCCATTCATCATATGAGGCTCATGCACTTACTATGATTTGATGACCAATAAA 710  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyValHisGlyLeuLeuAlaLeuVal 180  
DB 711 GTAGAAATTCCTTTTATGTTCTTTGATTTCTGGAGGTCACGTCGTGTGGCATTAAGT 770  
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyValSerLeuLeuAspIleAlaProGlyAspMet 200  
DB 771 CAAAGATTCACGATTTTCTGCTTCTTGAAAGCTCTTTGACATACACCAAGTGCATG 830  
QY 201 LeuAspLeuValAlaArgArgLeuSerLeuIleValHisProGluCysSerThrMetSer 220  
DB 831 CTGACAAAGTGGCAAGAAAGACTTCTTAAATAAAATCCAGAGTCTCCACATGAGT 890  
QY 221 GlyGlyValAlaIleGluHisLeuAlaValGlnGlyAsnArgPheHisPheAspIleVal 240  
DB 891 GGTGGGAAAGCCATGAAACATTTGGCCAAACAGAAATAGATTTTCAATTTTGACATCAA 950  
QY 241 ProProLeuHisHisAlaValAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 951 CTTCCCTTCATCAAGCTTAAATAATGATGATTTTCTTTTACTGGAGCTTCAACGCTTACT 1010  
QY 261 AspValIleIleMetLeuValSerGluGluGlyIleGluLeuSerGlyGlnIleLeuSer 280  
DB 1011 GATTAATAATATATGAAAAAGAAAAAGGAAAGATTTAGAAAGGGCAAAATCCGTGCT 1070  
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValIleValArg 300  
DB 1071 TCAGCAGCAGACATGCTGCCACAGTACAGCACACAAATGCAATGTCATCTTGGAAAGA 1130  
QY 301 ThrHisArgAlaIleLeuPheCysLeuSerGluArgAspLeuLeuProGlnMetAsnAlaVal 320  
DB 1131 ACACATCGGGCTAATTCCTGTTTGTAAAGCAGAGAGATTTGTACCTCAAAATATATGACATA 1190  
QY 321 LeuValAlaSerGlyValAlaSerAspMetLeuValAlaArgAlaLeuGluIleLeu 340  
DB 1191 CTGGTTCATCTGGTGGTGTCCCAAGTAATCTTATATCCGACAGCTCTGGAAATTTTA 1250  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
DB 1251 ACAACGCAACACAGTGCATCTTGTGTCTCTCCACAGCTATGACATGATATATGGC 1310  
QY 361 IleMetIleAlaTPanGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
DB 1311 ATTATGATTCATGAGTAATGATGATTAAGAAAGCTACGCTGGCTGGCAATTTTACATGAC 1370  
QY 381 IleGluGlyIleArgGlyArgProGlyCysProLeuGlyValAspIleSerGlyVal 400  
DB 1371 ATGAAAGCATCCGCTATGAAACCAAAATGTCCTCTTGAAGTACATATCAAAAGAACTT 1430  
QY 401 GlyGluAlaSerIleValProGluLeuLeuMetGluIle 414  
DB 1431 GGAGAAAGCTTCATMAAAGTACCAAAATTAATAATGAGATA 1472

RESULT 3  
AR428808 1387 bp DNA linear PAT 18-DEC-2003  
LOCUS AR428808 Sequence 21 from patent US 6642041.  
DEFINITION AR428808  
ACCESSION AR428808  
VERSION AR428808.1 GI:40186594  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1387)  
AUTHORS Chen, J., Feder, J. N., Nelson, T. C., Krystek, S. R. and Duclos, F.

TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;  
FEATURES Location/Qualifiers  
source 1..1387  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 3.54e-168 Length: 1387  
Score: 2090.50 Matches: 412  
Percent Similarity: 93.85% Conservative: 0  
Best Local Similarity: 93.85% Mismatches: 2  
Query Match: 98.38% Indels: 25  
DB: 6 Gaps: 1

US-10-649-273-2 (1-414) x AA428808 (1-1387)

QY 1 MetLeuIleLeuThrIleThraAlaGlyValPhePheIysProSerIleArgIleValIleTyr 20  
DB 24 ATGCTAATCTTGACTAAGACTGCAAGAGTTTATTTAAACATCAAAAGAAAGTTTAT 83  
QY 21 GluPheLeuArgSerPheAsnPhenHisProGlyThrLeuPheLeuHisIleValIleu 40  
DB 84 GAATTTTAAAGAAAGTTTAAATTTTTCATCTCGAACAACATATTCTTCATTAATAGTATTG 143  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60  
DB 144 GGAATTGAATAGTATGATGATACACAGCTGCTGCGTGATGAAGAACTGGAAATGTG 203  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuIleGlyIleValPro 80  
DB 204 TTGGAGAAAGCAATACATTTCCCAACTGAAGTTCAATTAAGACGGTGGAGTTGTCT 263  
QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB 264 CCAGCAGCTCAACAGCTTCAACAGAAATATTCAACGAATAGTCAAGAAAGCTTTCT 323  
QY 101 AlaSerGlyValSerProSerThrAspLeuSerAlaIleAlaThrThrIleIleProGlyLeu 120  
DB 324 GCCAGTGAAGCTCTCCCAAGTGAAGCTCTCAGCAATGTGCACTACATAAACCAAGACTT 383  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleuIleVal 140  
DB 384 GCTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACGCTGGTGAAGAGTTAAAG 443  
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160  
DB 444 CCATTCAATCCCATTCATCATATGAGAGCTCATGCACTTACATTAAGTTGACCATAA 503  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaVal 180  
DB 504 GTAGAAATTTCTTTTATTTAGTTCTTTGATTTCTTGAGAGTCACTGCTGTGGCATTAAGT 563  
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIleYssSerLeuAspIleAlaProGlyAspMet 200  
DB 564 CAAGAGAGTTTCAAGATTTCTGCTTCTTGAAAGTCTTTGAAATAGCAACCAAGTAACTAG 623  
QY 201 LeuAspIysValAlaArgArgLeuSerLeuIleIleHisProGluCysSerThrMetSer 220  
DB 624 CTTGACAAAGGTGGCAAGAGACTTTCTTTAATTAACATCCAGAGTGTCTCCACCATAGT 683  
QY 221 GlyGlyIysAlaIleGlnHisLeuAlaIysGlnGlyAsnArgPheHisPheAspIleIys 240  
DB 684 GGTGGAAAGCCATAGAGCATTTGGCCAAACAGAAATAGATTTCATTTGACATCAAA 743  
QY 241 ProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisIleValThr 260  
DB 744 CTTCTCTTGCAATGCTTAAATTTGTATTTTCTTTTACGTGACCTTCAACACGTTACT 803  
QY 261 AspIysIleIleMetIleValGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIle 273  
DB 804 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 863

QY 274 -----GluIys 275  
DB 864 CAGATAAATATTTCTGATGTGCTTAAAGCTCTCATTTCTGACGATAGGAAG 923  
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295  
DB 924 GGGCAATCTCTGTTCCAGCAGACAGCATTTGCTGCACAGTACAGCAACAATGCGATGT 983  
QY 296 HisLeuValValArgThrHisArgAlaIleLeuPheCysIysGlnArgAspLeuPro 315  
DB 984 CATCTTGAAAAGAACATCGGGCTATTCGTTTGTAGACAGAGACTGTGTACCT 1043  
QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPhenThrIleArgArg 335  
DB 1044 CAAATTAATGCAAGTCTGCTGATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1103  
QY 336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 355  
DB 1104 GCTTGAAATTTTAAACAAAGCAACAGTGCATTTGTGTGCTCTCTCCACAGACTA 1163  
QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375  
DB 1164 TCCACTGATTAATGGCATTATGATGATGATGATGATGATGATGATGATGATGATGATG 1223  
QY 376 GlyIleLeuHisAspIleGluGlyIleArgGlyIleArgGlyIleArgGlyIleArgGlyIle 395  
DB 1224 GGCATTTTACATGACATGAAAGCAGTCCGCTTGAACCAAAATGTCTCTTGAGATGAC 1283  
QY 396 IleSerIysGluValGlyIleGluAlaSerIleIleValProGlnLeuIleHisMetGluIle 414  
DB 1284 AATTCAAAGAAAGTTGGAAGACCTTCCATTAAGTACCAATTAATTAATGAAGATA 1340  
RESULT 4  
HSA295148 1387 bp mRNA linear PRI 30-OCT-2000  
LOCUS Homo sapiens mRNA for putative sialoglycoprotease type 2.  
DEFINITION  
ACCESSION AJ295148  
VERSION AJ295148.1 GI:11071726  
KEYWORDS metalloproteinase; sialoglycoprotease.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 Chen, J.M., Fortunato, M. and Barrett, A.J.  
Cloning and sequencing of a second human putative  
sialoglycoprotease homologue  
Unpublished  
2 (bases 1 to 1387)  
Chen, J.M.  
Direct Submission  
Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology  
Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,  
UNITED KINGDOM  
FEATURES  
source location/Qualifiers  
1..1387  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="uterus"  
24..1343  
/function="metalloproteinase of family M22"  
/codon\_start=1  
/product="putative sialoglycoprotease type 2"  
/protein\_id="CAC14666.1"  
/db\_xref="GI:11071727"  
/db\_xref="GOA:Q9H4B0"  
/db\_xref="UniProt/TREMBL:Q9H4B0"  
/translation="MULTITAGVFFKSKRYEFLRSPFNHPETLFLAKIVLGIET  
SCDDPAAVVDENGLVGLAISGTEVHLKGTGIVPAACQQLRENIQRIVEALASAS  
GVSPSDLSAIAITIKRGALSLGVGPSFLOLVGOLKRFPIIHHEAALITRLTNK  
VEPFLVLILSGHCLALVGVSPFLIGKSLDIAFGMDLKVARRSLIIRPECT

MSGGKALIEHLAKQGNRFHFDIKPRLHAKNCDPSFTGLQHVTKIMKKEBEGIFLI  
SKYBOINIPELCTKIAAHFCRYEKGOILSSADIAATVOHTMAKLVTRTHAILPCK  
ORDILPQNNNAVLSGGVAGSNFYIRBALEIETNATOCITLCPPLCTDNGIMLWANG  
IEHLRGLGILHIDIEGRYERKPCPLGVDISKEVGBASIKVQLKMEI\*

ORIGIN

Alignment Scores:  
Pred. No.: 3,54e-168 Length: 1387  
Score: 2090.50 Matches: 412  
Percent Similarity: 93.85% Conservative: 0  
Best Local Similarity: 93.85% Mismatch: 2  
Query Match: 98.38% Indels: 25  
Gaps: 1

US-10-649-273-2 (1-414) x HSA295148 (1-1387)

QY 1 MetLeuIleLeuThrLySerThraAlaGlyValIlePhePheLyProSerLyAspArgLyValIleYr 20  
DB 24 ATGCTAACTTGAAGTAAAGTCAAGAGAGTCTTTTAAACCATCAAAAAGAAAGCTTTAT 83  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLyValIleLeu 40  
DB 84 GAATTTTAAAGAGTTTAAATTTTCACTTGAACACTAATTTCTCATTAATAATAGTATTG 143  
QY 41 GlyIleGlyThrSerCysAspAspThrAlaAlaValIleAspGlyThrGlyAsnVal 60  
DB 144 GGAATTAAGTAACTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 203  
QY 61 LeuGlyGlyAlaIleHisSerGlyThrGlyValIleLeuLySerGlyValIleValPro 80  
DB 204 TTGGAGAGACCAATACATCCCAAGTGAATTCATTTAAACAGGAGGAGTCTTCT 263  
QY 81 ProAlaIleGlyGlyLeuHisArgGlyLeuAsnIleGlyValIleGlyValIleLeuSer 100  
DB 264 CCAGCAGCTCAACAGCTTCAAGAGAAATATTCACCAAAATAGTCAAGAAAGCTTTCT 323  
QY 101 AlAspGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLySerProGlyLeu 120  
DB 324 GCCAGTGAAGTCTTCCAGTGAAGTCTTCCAGTGAAGTCAATTCACCAATTAACAGAGCTT 383  
QY 121 AlaLeuSerLeuGlyValIleGlyLeuSerPheSerLeuGlyValIleGlyValIleLeuSer 140  
DB 384 GCTTTAAGCCTGGAGAGGCTTATCATTTAGCTTACAGCTGATGAGACAGTAAAGAAAG 443  
QY 141 ProPheIleProIleHisSerGlyAlaHisAlaLeuThrIleArgLeuThrAsnLy 160  
DB 444 CCAATTCATTCCTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 503  
QY 161 ValGluPheProPheLeuValIleLeuIleSerGlyValIleCysLeuLeuAlaLeuVal 180  
DB 504 GTAGAAATTTCTTTTATGATTTCTTGAAGTCACTGCTGTGCTGCTGCTGCTGCTGCT 563  
QY 181 GluGlyValSerAspPheLeuLeuGlyLySerSerLeuAspIleAlaProGlyAspMet 200  
DB 564 CAAGAGATTTCAAGATTTCTGCTTCTGGAAGCTTTTGGACATGACACAGTACATG 623  
QY 201 LeuAspLyValAlaArgArgLeuSerLeuIleLyHisProGlyCysSerThrMetSer 220  
DB 624 CTTCACAGAGGAGGAGAGAGCTTTCTTAAATTAACATCCAGAGTGTCCACCATGAGT 683  
QY 221 GlyGlyLyValAlaIleGlyHisLeuAlaLySerGlyValAsnArgPheHisPheAspIleLy 240  
DB 684 GTGGGAGAAACCAATGAGCATTTGGCCAAACAGAAATAGATTCATTTTGAACATCAAA 743  
QY 241 ProProLeuHisHisAlaLyAsnCysAspPheSerPheThrGlyLeuGlyHisValThr 260  
DB 744 CTTCCCTTGCACTACGCTTAAATTTGATTTTCTTTTACTGACCTTCACACACCTTACT 803  
QY 261 AspLyIleIleMetLyGlyGlyValIleGlyValIleGlyValIleGlyValIleGlyVal 273  
DB 804 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 863  
QY 274 -----GlyLyS 275

DB 864 CAGATAATATTTCTGATTTGCTTAAATAGCTGCTCATTTCTGACGTATGAGAG 923  
QY 276 GlyGlyIleLeuSerSerAlaAlaAspIleAlaIleValGlyHisThrMetAlaCys 295  
DB 924 GAGCAATCTGCTTCCAGACAGACATTCCTGACAGATGACACAAATGACATGT 983  
QY 296 HisLeuValLyArgThrHisArgAlaIleLeuPheCysLySerGlyAspLeuPro 315  
DB 984 CATCTTGTAAGAAAGACATCGGCTATTCCTTTGTAAGAGAGAGCTTGTACT 1043  
QY 316 GlnAsnAsnAlaValIleValAlaSerGlyValIleAlaSerAsnPheThrIleArg 335  
DB 1044 CAAATTAATGACGATCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1103  
QY 336 AlaLeuGlyIleLeuThrAsnAlaThrGlyCysThrLeuLeuCysProProAlaGlyLeu 355  
DB 1104 GCTTGGAATTTTAAACAAACCAACAGTGCATCTTGTGCTCTCTCCACAGCTA 1163  
QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGlyArgLeuArgAlaGlyLeu 375  
DB 1164 TGCATGATTAATGCTTATGATGATGATGATGATGATGATGATGATGATGATG 1223  
QY 376 GlyIleLeuHisAspIleGlyValIleArgGlyValProLySerProLeuGlyValAsp 395  
DB 1224 GGCATTTTACATGACATAGAGGACCGCTATGACCAAAATGTCCTTGAGATGAC 1283  
QY 396 HisSerLyGlyValIleGlyValIleAspIleLyValProGlyLeuLySerGlyIle 414  
DB 1284 ATATCAAAAGAGTGGAGAGCTTCCATTAAGTCACAATTAATAATGAGATA 1340  
RESULT 5  
AX664697 1245 bp DNA linear PAT 22-MAR-2003  
LOCUS AX664697 Sequence 6 from Patent WO02074960.  
DEFINITION AX664697  
ACCESSION AX664697  
VERSION AX664697.1 GI:29164457  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
Leiby, K.R., Kapeller-Libermann, R. and Gluckmann, M.  
38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
of human proteins and uses thereof  
Patent: WO 02074960-A 6 26-SEP-2002;  
JOURNAL Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
Location/Qualifiers  
source 1..1245  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 5.03e-168 Length: 1245  
Score: 2088.00 Matches: 407  
Percent Similarity: 99.03% Conservative: 3  
Best Local Similarity: 98.31% Mismatches: 4  
Query Match: 98.26% Indels: 0  
Gaps: 0  
US-10-649-273-2 (1-414) x AX664697 (1-1245)  
QY 1 MetLeuIleLeuThrLySerThraAlaGlyValIlePhePheLyProSerLyAspArgLyValIleYr 20  
DB 1 ATGCTAACTTGAAGTAAAGTCAAGAGAGTCTTTTAAACCATCAAAAAGAAAGCTTTAT 60  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLyValIleLeu 40  
DB 61 GAATTTTAAAGAGTTTAAATTTTCACTTGAACACTAATTTCTCATTAATAATAGTATTG 120



```

Qy 41 GlyIleGIuThrSerCyAspAspThrAlaAlaValAlaAspGIuThrGIyAsnVal 60
Db 121 GGAAATTGAACTAGTTGATGATGATACAGACAGCTGCTGCTGATGATGAACCTGGAATGTG 180
Qy 61 LeuGIyGIuAlaIleHisSerGIuThrGIuValHisLeuLySerThrGIyGIyIleValPro 80
Db 181 TTGGAGAGGAACAATACATTTCCCAACTGAACTCATTTTAAAAACAGTGGGATGTTCTCT 240
Qy 81 ProAlaAlaGIuThrSerCyAspAspThrAlaAlaValAlaAspGIuThrGIyAsnVal 100
Db 241 CCAGAGGCTCAACAGCTTCAACAGAAATATTCACCAATATGATCAAGAACTCTTCTCT 300
Qy 101 AlaSerGIyValSerProSerAspLeuSerAlaIleAlaThrThrIleValProGIyLeu 120
Db 301 GCCAGTGAAGCTCTCCCAAGTACCTCTCAGCAATGTGACATACCAATAAACAGAGACTT 360
Qy 121 AlaLeuSerLeuGIyValGIyLeuSerPheSerLeuGIuValGIyGIuThrGIyLeu 140
Db 361 GCTTTAAGCTGGGAGGCTGATCATTTAGCTTACAGCTGTGTGAGCAGTTAAAAAAG 420
Qy 141 ProPheIleProIleHisIleMetGIuAlaHisAlaAlaThrThrIleArgLeuThrAsn 160
Db 421 CCATTCAATCCCATTCATCATATGAGAGGCTCATGCACTTACTATTAGTTGACCAATATA 480
Qy 161 ValGIuPheProPheLeuValLeuLeuIleSerGIyGIyHisCyAspLeuValAlaLeuVal 180
Db 481 GTAGAAATTTCTTTTATGATCTTTTGAATTTCTGAGAGTCACTGCTGTGGCATTTAGT 540
Qy 181 GIuGIyValSerAspPheLeuLeuLeuGIyLySerLeuAspIleAlaProGIyAspMet 200
Db 541 CAAGAGATTTCAAGATTTTCTGCTTCTGAAAGCTTTTGAGCATATGACACAGTGCATG 600
Qy 201 LeuAspLyValAlaArgArgLeuSerIleuIleValHisIleProGIuCySerThrMetSer 220
Db 601 CTTCACAGAGGCGGCAAGAGACTTTCTTTAATAAACATCCAGAGTGTCCCATGTAGCT 660
Qy 221 GIyGIyValAlaIleGIuHisIleuAlaLySGInGIyAsnArgPheHisIleAspIleVal 240
Db 661 GGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTTCAATTTTACATCCAA 720
Qy 241 ProProLeuHisIleAlaIleValAlaThrValGIuHisIleThrMetAlaCyHisIleuValIyAsn 260
Db 721 CCTCCCTTGCATCAGCTGATGCTGCAAGTACACACATGCGCATCTTGTGAAGAA 900
Qy 301 ThrHisArgAlaIleLeuPheCySerLySGInArgAspLeuLeuProGIuHisAsnAlaVal 320
Db 901 ACACATGGGGCTATCTCTTTTGTAGCAGAGAGACTTGTATCTCAAAATATATGACAGA 960
Qy 321 LeuValAlaSerGIyGIyValAlaSerAsnPheIleValArgAlaLeuGIuIleVal 340
Db 961 CTGGTTCATCTGGGCTGCTGCAAGTAACTTCTATCCGACAGCTCTGCAAAATTTTAA 1020
Qy 341 ThrAsnAlaThrGIuThrLeuLeuCySerProArgArgLeuCySerThrAsnArg 360
Db 1021 ACAAACGCAACACAGTGCATTTGTTGTCTCTCCACAGCTATGCACTATTAATGTC 1080
Qy 361 IleMetIleAlaIleProAsnGIyIleGIuArgLeuArgAlaGIyLeuGIyIleLeuHisAsp 380
Db 1081 ATTATGATTCAGTGAATGATGATTAAGAAAGCTACCTGCTGTGGCATTTTACATGAC 1140
Qy 381 IleGIuGIyIleArgIyGIuProLyCyProLeuGIyValAspIleSerLyGIuVal 400
Db 1141 ATAGAAGGCAATCCGCTATGAACCAAAATGTCTTGGAGTGAATATCAAAAGAGTT 1200
Qy 401 GIyGIuAlaSerIleValValProGIuLeuLyMetGIuIle 414

```

```

Db 1201 GGAGAGCTTCCATAAAGTACCAACAATTAAGAGATGATA 1242
RESULT 6
AX664695 1820 bp DNA linear PAT 22-MAR-2003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
Leiby, K.R., Kessler-Libermann, R. and Gluckmann, M.
38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions
of human proteins and uses thereof
Patent: WO 02074960-A 4 26-SEP-2002
JOURNAL
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1..1820
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
146..1390
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD80044.1"
/db_xref="GI:29164456"
/translation="MULTITAGVPEPKSKRYEFLRSFNFPGTLPLKIVLGLET
SCDDPRAAVVDSTGNVIGRAIHSGTBEVHLKTGGIYPPAQQOLHREHIOIVQALAS
GYSBPDLSAINTTIRPGALSLGVSLSLQYGLKRPFIHHHEAALTRITLNK
VEFPVLVLTISGHCILALVGVSPDLTKSLDLPGLMDKVAERLILKIPBST
MSGRAIEHLAQGNRFPDIPPLHAKNCPSTGLQHTVDKNNENKQESIEKG
QILSSAADIATVQHTMACHLVETRRAILPKORDLLEPNNAAVVAAGVSNFYIR
RLRLITMAQCTCLCPRLCTDNGIMAMNGIBRLRAGLGIHLDIEGIRYEPKCP
GVDISKEYGEASIKVPLKMET"
ORIGIN
Alignment Scores:
Pred. No.: 8.14e-168 Length: 1820
Score: 2088.00 Matches: 407
Percent Similarity: 99.03% Conservative: 3
Best Local Similarity: 98.31% Mismatches: 4
Query Match: 98.26% Indels: 0
DB: Gaps: 0
US-10-649-273-2 (1-414) x AX664695 (1-1820)
Qy 1 MetLeuIleLeuThrIleThrAlaGIyValPhePheLyProSerLyAsnArgLyValIle 20
Db 146 ATGCTAATCTTCACTAAGACCTCAGAGAGTTTAAAAACATCAAAAGAAAGTTAT 205
Qy 21 GIuPheLeuArgSerPheAsnPheHisIleProGIyThrLeuPheLeuHisIleValIleu 40
Db 206 GAATTTTAAAGAGTTTAAATTTTCACTCTGGAACCTATTTCTTATTAATAATGATATG 265
Qy 41 GlyIleGIuThrSerCyAspAspThrAlaAlaValAlaAspGIuThrGIyAsnVal 60
Db 266 GGAAATTGAACTAGTTGATGATGATACAGCAGCTGCTGTGTGATGAATGAATGTG 325
Qy 61 LeuGIyGIuAlaIleHisSerGIuThrGIuValHisIleuLySerThrGIyGIyIleValPro 80
Db 326 TTGGAGAGCAATACATTTCCAAACTGAAGTTCAATTTAAAAACAGTGGGATGTTCTCT 385
Qy 81 ProAlaAlaGIuThrLeuHisArgGIuAsnIleGIuArgIleValGIuGIuAlaLeuSer 100
Db 386 CCAGAGCTCAACAGCTTCAACAGAAATATTCACCAAGATGATACAGAAAGCTTTTCT 445
Qy 101 AlaSerGIyValSerProSerAspLeuSerAlaIleAlaThrThrIleLyProGIyLeu 120
Db 446 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTTGCAACTACATTAACACAGACTT 505

```

QY 121 AAlaSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuValys 140  
DB 506 GCTTAAAGCTGGAGGAGGCTTATCATTTAGCTACAGCTGTAAGAGAGTTAAAG 565  
QY 141 ProPheLeuProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAnlys 160  
DB 566 CCATTTCATTCCTCATTCATATGAGGCTCATGCACTTACTATTAGGTTGACATTA 625  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyHisCysLeuLeuAlaLeuVal 180  
DB 626 GTAAATTTCTTTTATTTAGTTCTTTGANTTCTGAGGCTGCTGTGTGACATTAGT 685  
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
DB 686 CAAGAGTTTCAGATTTTCTGCTTCTTGAAAGCTTTTGACATACAGCAAGTACATG 745  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
DB 746 CTTGACAAAGGTGGCAAGAAAGCTTTCTTAATTAACATCCAGAGTCTCCACATGAGT 805  
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAspArgPheHisPheAspIleLys 240  
DB 806 GGTGGAAAGCCATTAAGAAACATTTGGCCAAACAAAGAAATAGATTTCATTTGACATCAA 865  
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 866 CTTCCCTTGCATCATCTGCTAAATTTGATTTTCTTTTACCTGAGCTTCAACACCTTACT 925  
QY 261 AspLysIleIleMetLysLysGlnLysGlnGlyIleGluLysGlyGlnIleLeuSer 280  
DB 926 GATTAATAATATAGAAACAGAAACAAAGAAAGATTTGAGAAAGGGCAATCTGCT 985  
QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisSerMetAlaCysHisLeuValLysArg 300  
DB 986 TCAGCACAGACATGCTGCAAGTACAGACACAAAGGCAATGCTATCTGTGAAAAGA 1045  
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
DB 1046 ACACATGGGGCTATTCGTTTGTGTAAGCAAGAGCTGTGTAACCTCAAAATATATCAGTA 1105  
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340  
DB 1106 CTGTTTCATCTGCTGCTGCTGCAAGTAACTTCTATATCCGAGAGCTCTGGAATTTTA 1165  
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgGluCysThrAspAsnGly 360  
DB 1166 ACAAAACCAACACAGTGCACTTGTGTGCTCTCCCAAGCTATGCACTGATATAGGC 1225  
QY 361 IleMetIleAlaTyrAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
DB 1226 ATTATGATTCATGTAAGATGTAATGAAAGACTACCTGCTGGGCAATTTTATATAC 1285  
QY 381 IleGlnGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGlnVal 400  
DB 1286 ATAGAAGGCATCCCTATGAACCAAAATGCTCTTGAAGTACATATCAAAAGAGTT 1345  
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
DB 1346 GGAAGAGCTTCATTAAGATGACCAATTAATAATGAGATTA 1387  
RESULT 7  
AX713716 2208 bp DNA linear PAT 15-APR-2003  
LOCUS Sequence 400 from Patent EP1293569.  
DEFINITION AX713716  
ACCESSION AX713716  
VERSION AX713716.1 GI:29888642  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,  
Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,  
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and  
Masuko, Y.  
TITLE Full-length cDNAs  
JOURNAL Patent: EP 1293569-A 400 19-MAR-2003;  
Helix Research Institute (Jp) ; Research Association for  
Biotechnology (Jp)  
FEATURES  
source Location/Qualifiers  
1..2208  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,83e-155 Length: 2208  
Score: 1944.00 Matches: 386  
Percent Similarity: 93.24% Conservative: 0  
Best Local Similarity: 93.24% Mismatches: 4  
Query Match: 91.48% Indels: 24  
DB: Gaps: 1  
US-10-649-273-2 (1-414) x AX713716 (1-2208)  
QY 1 MetIleuIleLeuThrIleThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
DB 344 ATGCTATATCTTGACTTAAGAGCTGCAAGAGTTTATTTTAAACATCAAAAGAAAGTTAT 403  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
DB 404 GATTTTAAAGAAAGTTTATTTATTTTCATCTGGAACACTATTTCTTCAATAAATAGTATTG 463  
QY 41 GlyIleGlnThrSerCysAspAspThrAlaAlaValAlaLysGlnThrGlyLysAsnVal 60  
DB 464 GGAATTAAGAACTAGTTGTATGATATACAGACAGCTGCTGTGTGATGAAACTGGAATATGTG 523  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIleValPro 80  
DB 524 TTGGGAGAAAGCAATATATTTCCCAAGTGAAGTTCAATTTAAACAGGTGGAGATTTGCT 583  
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB 584 CAGAGCTCAACAGCTTCAACAGAAATATTCAAGANTATATACAGAGCTCTTTCT 643  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
DB 644 GCAGTGGAGTCTCTCCAAAGTGACTCTCAGCAATTTGCAACTACATTAATAACAGAGACTT 703  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
DB 704 GCTTTAAAGCTGGAGTGGCTTATATATTAGCTTACAGTGTATGACAGTTAAAAAG 763  
QY 141 ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAnlys 160  
DB 764 CCATTTCATTCCTCATTCATATGAGGCTCAAGCACTTACTATTAGTTGACCAATTA 823  
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 824 GTAGAAATTTCTTTTATTTAGTTCTTTGATTTTCGAGAGTCACTGCTGTGGCATTAAGTT 883  
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
DB 884 CAAGAGTTTCAGATTTTCTGCTTCTTGAAAGCTTTTGACATACACCAAGGTGACATG 943  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
DB 944 CTTGACAAAGGTGGCAAGAAAGCTTCTTAATTAACATCCAGAGTCTCCACAGATGACT 1003  
QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAspArgPheHisPheAspIleLys 240  
DB 1004 GGTGGAAAGCCATTAAGAAACATTTGGCCAAACAAAGAAATAGATTTCATTTGACATCAA 1063

[illegible]

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Toogishi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shitai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuno, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat. Genet.* 36 (1), 40-45 (2004)

2  
 Nishi, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Ito, R., Otsubi, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai, Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Waga, T., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuno, Y., Nagai, K. and Isogai, T. NEDO human cDNA sequencing project

Unpublished  
 3 (bases 1 to 2208)  
 Isogai, T., Otsubi, T. and Sugiyama, T.  
 Direct Submission  
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 153-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB) (supported by Japan Construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.), 5'-5'-3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES  
 source  
 1. 2208  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="FEBRA2004592"  
 /issue\_type="brain"  
 /clone\_lib="FEBRA2"  
 /dev\_stage="fetus"  
 /note="cloning vector: pME18SFL3"  
 344. 1438  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB70923.1"  
 /db\_xref="GI:16550167"  
 /translation="MLLTITKTAQVFFKPSKRYEFLRSNFPHPGTLFLKIVLGIEITSCDDTAAAVDDENGVIGEAHSGQTEVHLKTGSI VPPAAQQLRENIQRIQVQALSSGSPSDLSAATTIKPGALISLGVLSFQLQGVQAKKPIPIHMEAHALITRLTKVVEPPVILISGGHLLAQVGSDFLLGKSLDIAVGMDLVAARLPIIKKESTMSGKALIEHLAKQGNFPIKPIKPLHNAKQCDPPGLOHVDKTIIMKEKEGIENGQIISSAADIAATIQHTMACHLVKTRAILIFCQRDLIPQNAVLVASGVASNF CIRRALEILMTAQCTLLCPRLCTDNGIMIA"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.83e-155 Length: 2208  
 Score: 1944.00 Matches: 386  
 Percent Similarity: 93.24% Conservative: 0  
 Best Local Similarity: 93.24% Mismatches: 4  
 Query Match: 91.48% Indels: 24  
 DB: 9 Gaps: 1

US-10-649-273-2 (1-414) x AK055441 (1-2208)

QY 1 MetheulleuethrlyrthrralaglyvalphelphelysProserlylsArglysvaItyr 20  
 Db 344 ATGTATATCTTGAACTAAGACTGCAGGAGCTTTTTTTTAAACATCAAAAGAAAGTTAT 403  
 QY 21 GluPhelauarGserPheanPhehIsPfcGlyThrlenuPhelauhIslysvIlevalleu 40

```

Db      404 GAATTTTAAAGAGTTTAAATTTTCATCTCGAGACATATTTCTCATAAAAATGATATTG 463
Qy      41  G1Y1leGIuThrSerCyAspAspThrAlaAlaValAlaAspGIuThrGIYAsnVal 60
Db      464 GGAATTGAAATAGTATGATGATATACACAGCTGCTGCTGCGATGAAATCGAATGCG 523
Qy      61  LeuGIYGIuAlaIleHISerGIuThrGIuValHISleuYsThrGIYGIYIleValPro 80
Db      524 TTGGAGAGAGCAATACATTTCCCAACTGAGATTCATTTAAAAACAGTGGATGTTCTCT 583
Qy      81  ProHlaIaGIuInleuHISerGIuAsnIleGIuValIleValGIuInleuAlaIleuSer 100
Db      584 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATAGTCACAAAGCTCTTCTCT 643
Qy      101 AlAserGIYAlaSerProSerAspSerAspSerAlaIleAlaThThrIleYsProGIYleu 120
Db      644 GCCAGTGAAGTCTCTCCAGAGTACCTCTCAGCAATTTGCACTTAAACATTAACAGGACTT 703
Qy      121 AlAlaSerleuGIYAlaGIYleuSerPheSerleuGIuValGIYGIuInleuYsIys 140
Db      704 GCTTTAAGCCTGGAGTGGCTTATTCATTTAGCTTACAGCTGTAAGACAGTTAAAAAAG 763
Qy      141 ProPheIleProIleHISerMetGIuAlaHISAlaIleuThrIleArgleuThraAsnIys 160
Db      764 CCATTCAATCCCATTCATCATATGAGGCTCATGCACTTACTATTAGGTTGACCAATATAA 823
Qy      161 ValGIuTherProPheleuValleuIleSerGIYGIYHISCyaleuIleuAlaIleuVal 180
Db      824 GTAAATATTTCTTTTATAGTCTTTGATTTCTGAGAGTCACTGCTGTTGGCATTAGTT 883
Qy      181 GIuInGIYAlaSerAspPheleuIleuGIYIysSerleuAspIleAlaProGIYAspMet 200
Db      884 CAAGAGATTCAGATTTTCTGCTCTTTCGAAAGCTTTTGACATAGCACAGTATGACATG 943
Qy      201 LeuAspIysValAlaArgArgleuSerleuIleYsHISProGIuCySerThrMetSer 220
Db      944 CTTGACAAAGGCGCAAGAGACTTCCTTAATAAAACATCCAGAGTCTCCACATGAGT 1003
Qy      221 GIYGIYIysAlaIleGIuHISleuAlaIysGIuInleuValAsnArgPheAspIleYs 240
Db      1004 GGTGGGAAGCCATAGAACATTTGGCCAAACAGAAATAGATTCATTTGACATCAAAA 1063
Qy      241 ProProleuHISHISAlaIysAsnCyAspPheSerPheThrGIYleuGIuHISValThr 260
Db      1064 CTTCCCTTCGATCATGCTTAAATTTGATTTTCTTTTACGAGCTTCACACAGCTTACT 1123
Qy      261 AspIysIleIleMetIysIysGIuIysGIuGIuGIYIleGIuIysGIYGIuInleuSer 280
Db      1124 GATTAATTAATTAATGAAGAAAGAAAGAGAGAGTATTTGAAGAGGGCAAAATCTGCTCT 1183
Qy      281 SerAlaAlaAspIleAlaIleThValGIuHISThrMetAlaCyHISleuValIysArg 300
Db      1184 TCAGCACAGACATGCTGCACAGTACAGACACAAATGCAATGCACTTCTGTGAAGA 1243
Qy      301 ThrHISArgAlaIleleuPheCyaleuGIuArgAspIleuProGIuAsnAsnAlaVal 320
Db      1244 ACACATGGGGCTATTCGTTTGTGAAGAGAGAGCTTTGATCCCAAAATATATCAGTA 1303
Qy      321 LeuValAlaSerGIYGIYAlaAlaSerAsnPheThrIleArgArgAlaIleuGIuIleu 340
Db      1304 CTGTTTGCATTCGTGCTGCTGCAAGTACTTCTGATCCGACAGCTCTGGAATTTTA 1363
Qy      341 ThrAsnAlaThrGIuYsThrleuIleuCyAspProProArgleuCyThraAsnArgIy 360
Db      1364 ACAAAACCAACAGTGCATCTTGTGTGCTCTCTCCACAGATATGACATATATATGCGC 1423
Qy      361 IleMetIleAlaThraAsnArgIyIleGIuArgleuArgAlaGIYleuGIYIleleuHISAsp 380
Db      1424 ATTAATGATGCA-----TGATGCTCTTGGAGTACATATCAAAAGAGTT 1435
Qy      381 IleGIuGIYIleArgIyArgIuProIysCyAspProleuGIYAlaAspIleSerIysGIuVal 400
Db      1436 -----TGATGCTCTTGGAGTACATATCAAAAGAGTT 1471

```

```

Qy      401 G1Y1leGIuAlaSerIleYsValProGIuInleuYsMetGIuIle 414
Db      1472 GGAAGAGCTTCCATTAAGTACACATTAATAATGAGAGTA 1513

RESULT 9
BC058172
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCES
AUTHORS
JOURNAL

REMARK
COMMENT

NIH-MGC Project URL: http://mgc.nci.nih.gov
Contract: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akheri, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghghi, P.,
Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Latic, P., Legaspi, R.,
Maduro, Q.L., Masello, C., Maskeri, B., Mastriani, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantrijop, S., Thomas, P.J., Touchman, J.W.,
Taugenon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAX Plate: 123 Row: d Column: 22
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21312463.
Location/Qualifiers

```

source

1. 1844  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="MGC:67870 IMAG:5012054"  
/issue\_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."  
/clone\_id="NCI CGAP\_1429"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
222. 1466  
/codon\_start=1  
/product="Unknown (protein for MGC:67870)"  
/protein\_id="AAH58172.1"  
/db\_xref="GI:34849664"  
/translation="MLMLRTYAGAIIPKPKSKYVGLRFRSVHPRTLSCHKLVIGIET  
SCDPTGAIVDETVGVGALHSQTVHLKKGIVPPVAQQLHEHNIQRIYETLSAS  
RTPSDLSAIVTKIPGLASIGVLSLQVNRFKKPIPIHMEHALTIRLTKT  
VEPFIYLIISGCHLAVGVSDPILGLSLDAPDMIDKVARLSLKHPCST  
MSGKALHIAKDKRRHFTINPPMNKAKNDFSTGICHTDKLITKEKEBIEKG  
QILSSADIVAAVQAHATCAKRTTHRAILFCQKNLSPNAAVLVSGVAASNLVIR  
KALEIVANATQCTLLCPPLCTDNGIMIANNGIERLRAGLVLDVEDIYEPKPL  
GIDISREVAEAIVPRLKMAI"

misc\_feature  
333. 1397  
/note="ORF17, Region: Metal-dependent proteases with possible chaperone activity [posttranslational modification, protein turnover, chaperones]"  
/db\_xref="CDD:COG0533"

ORIGIN

Alignment Scores:  
Pred. No.: 2.71e-146 Length: 1844  
Score: 1835.00 Matches: 352  
Percent Similarity: 91.06% Conservative: 25  
Best Local Similarity: 85.02% Mismatches: 37  
Query Match: 86.35% Indels: 0  
DB: 10 Gaps: 0

US-10-649-273-2 (1-414) x BC058172 (1-1844)

QY 1 MetLeuIleLeuThrValAGlyValPhePheProSerLysArgLysValTyr 20  
DB 222 ATGCTAATGTTAAGAAGACAGCAGACCTATCCCAAGCCCAAGTAAGTTAT 281  
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLeuValLeu 40  
DB 282 GGAATTTTAAAGAGATTAGTTCATCCAGAACTCTCTGTCATTAACCTGGTCCG 341  
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyAsnVal 60  
DB 342 GGAATTTAAGACACTGTGATGACACAGAGCGCTGTGTGTGATGAAGAACTGGGATGTG 401  
QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyValLeuPro 80  
DB 402 CTGGGGAGAGACGTGCACTCCCAACTCAGGTTCACTGAAACAGGTGGATTTTCT 461  
QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB 462 CCAAGTAGCTCAACAATTCACAGAGAAATATTCACAGAAATAGTAGAAGAACTTTTCT 521  
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
DB 522 GCCAGTAGAATCACCCCAAGCATCTCTGGCAATTTGCAACTACATCAACCGGAGCTG 581  
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140  
DB 582 GCCCTAAGCTTAGAGAGTGGCTTATCTTTAGCTTACAGCTAGTAATTAATTTAAAG 641  
QY 141 ProPheIleProIleHisIleMetGlyValHisAlaLeuThrIleArgLeuThrAsnLys 160  
DB 642 CCGTTCATCCGATTCATCACTGAGAGCTCAGCATGACTATTAAGCTCACCAATATA 701

QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
DB 702 GTAGAATTTCTTTCTTCTAGTTCTTTGATTTCTGGGGGTCACTGCTGTGTGGATTAAGTC 761  
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
DB 762 CAAGGTGTTCCGATTTCCCTGCTCTGGAGAGTCTTTGACATAGCCGAGGCGACAG 821  
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
DB 822 CTGACCAAGGTGGCAAGAGACTTTCTTTAATCAACATCCGAAATGTTCTCAATGAGT 881  
QY 221 GlyGlyLysAlaIleGlyHisLeuAlaLysGlnGlyAsnArgPheHisAspAspIleLys 240  
DB 882 GGTGAAAAGCTATGAACTTTGGCCAAAGCAGAAATGATTCATTTACTATCAAT 941  
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB 942 CCACCTATGCAAGATGCTAAGATTTGCGATTTTCTTTACGGGACTTCAACATATTACT 1001  
QY 261 AspLysIleIleMetLysLeuGlyLysGluGlyIleGlnLysGlyGlnIleLeuSer 280  
DB 1002 GATTAAGCTAATTAACAACAAGAGAAAAGAGAGCATTTGAGAGAGGCGAAATCCCTGCA 1061  
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
DB 1062 TTAGCTGACAGATTTCTGCTCGGTACAGATTCACAGCCGTGCACTTGCAGAAAG 1121  
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaVal 320  
DB 1122 ACACATCGTCAATTCCTGTTTGGCAAGCAAGAAAATTTGCTATCTCCAGCTAACGAGTA 1181  
QY 321 LeuValAlaAspGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeu 340  
DB 1182 TTAGTGTATCTGAGAGTGTTCAGATTACTGTATCCAGAAAGCATTTGAAATGTGTC 1241  
QY 341 ThrAsnAlaThrGluCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
DB 1242 GCAATAGCAAGCAGAGGACGTGCTGTGTCCTCCCAAGACTGTGACATGACAAATGGC 1301  
QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
DB 1302 ATCATGATTTGCATGGAATGGAATGGAATTAACGTACGTGGCGCTTTTACATGAT 1361  
QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
DB 1362 GTAGAAGCATCCGATATGAAACCAAAATGCTCTTGGAAATGACATATCCAGAGAAATT 1421  
QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
DB 1422 GCAAGAGCTGCATTAAGATCCGCAATTAAAGTGAAGCACTT 1463

RESULT 10  
ARS41929 1416 bp DNA linear PAT 08-OCT-2004  
LOCUS DEFINITION  
SEQUENCE 177 from patent US 6743619.  
ACCESSION ARS41929  
VERSION ARS41929.1 GI:53934009  
KEYWORDS  
SOURCE  
ORGANISM  
Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 1416)  
Tang, Y.T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F.,  
Zhang, J., Zhao, Q.A., Yang, Y., Xue, A.J., Wehrman, T., Wang, J.-R.,  
Wang, D. and Dmanac, R.T.  
Nucleic acids and polypeptides  
Patent: US 6743619-A 177 01-JUN-2004;  
location/Qualifiers  
1. 1416  
/organism="unknown"  
/mol\_type="genomic DNA"

TITLE  
JOURNAL  
FEATURES  
source

## ORIGIN

## Alignment Scores:

Pred. No.:	5.9e-139	Length:	1416
Score:	1747.00	Matches:	340
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.71%	Mismatches:	0
Query Match:	82.21%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2 (1-414) x AR541929 (1-1416)

```

QY 74 LyeThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93
DB 280 AGAAGAGGTGGAGATTGTTCTCCAGAGCTCAACAGCTTCAAGAGAAATATTCAACGA 339
QY 94 IleValGlnGlnValAlaLeuSerAlaSerGlyValSerProSerArgPheSerAlaIleAla 113
DB 340 ATAGTACMAAGAGCTTTCTGCGAGAGAGTCTTCCAACTGACCTTCAGCAATTGCA 399
QY 114 ThrThrIleLeuProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
DB 400 ACTACCACTAAACAGAGCTTCTTAAAGCTGGAGTGGGCTTATTCATTAGCTTACAG 459
QY 134 LeuValGlyGlnLeuValysPheProPheIleProIleHisIleMetGluAlaHisAlaLeu 153
DB 460 CTGGTAGAGACAGTTAAAGAGCCATTCTCCATTCTATCATATGAGAGCTCATGCACTT 519
QY 154 ThrIleArgLeuThrAsnValGlnPheProPheLeuValLeuLeuIleSerGlyGly 173
DB 520 ACTATTAGGTTGACCAATTAAGTGAATTTCTTTTATGTTCTTTGATTCTTGAGAGT 579
QY 174 HisCysLeuLeuAlaLeuValGlnGlyValSerArgPheLeuLeuGlyValysSerLeu 193
DB 580 CACTGCTGTGTGATTAGTTCAAGAGGAGTTTCAATTTTCTGCTTGGAAAGCTTTTG 639
QY 194 AspIleAlaProGlyAspMetLeuAspValAlaArgArgLeuSerLeuIleHis 213
DB 640 GACATGACACAGGTGACATGCTTGAAGAGTGGCAAGACATTTCTTTAATTAAACAT 699
QY 214 ProGluCysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaValysGlnGlyAsn 233
DB 700 CCAGAGGTCTCCACATGAGTGGTGGAGAACCATAGAACATTTGGCCAAACAGAAAT 759
QY 234 ArgPheHisPheAspIleLeuProProLeuHisIleAlaValysAsnCysAspPheSerPhe 253
DB 760 AGATTTCATTTTGCATCAACCAACCTCCCTTCATCATGCTTAAAAATTTGATTTTCTTT 819
QY 254 ThrGlyLeuGlnHisValThrAspValIleIleMetValysGluValysGluGlyIle 273
DB 820 ACTGACCTTCAACGCTTACTGATTAATAATTAATTAAGAAAGGAAAGAGGATAT 879
QY 274 GluValysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMet 293
DB 880 GAGAAAGGGGCAATTCCTGCTTCAAGAGCAAGACTGCTGCACAGTACAGCAACATG 939
QY 294 AlaCysHisLeuValValysArgThrHisArgAlaIleLeuPheCysValysGlnArgAspLeu 313
DB 940 GCATGTCATCTTGTGAAGAAACACATCGGGCTATTCTGTTTGAAGACAGAGACTTG 999
QY 314 LeuProGlnAsnAsnAlaValIleLeuValAlaSerGlyValAlaAlaSerAsnPheTyrIle 333
DB 1000 TTACCTCAAAATATATGATGATGCTGTCATCTGTGTGTGCGAAGTACCTTCATATATC 1059
QY 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro 353
DB 1060 CGAGAGCTCTGAAATTTTAAACAGCAACAGTGCACCTTGTGTGTCTCTCC 1119
QY 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAla 373
DB 1120 AGACTATGACCTGATATATGCAATTATGATGCAATGATATTTGAAGACTACGTCT 1179
QY 374 GlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGly 393

```

```

DB 1180 GCGTTGGGCACTTTACATGACATAGAGGATCCGCTATGAAACCAAAATGCTCTTGG 1239
QY 394 ValAspIleSerValGlnValGlyValAlaSerIleLeuValProGlnLeuValMetGln 413
DB 1240 GTAGACATATTCAAAAGAAAGTTGAGAGAGCTTCATAAAGTACCAACATTTAAATGAG 1299
QY 414 Ile 414
DB 1300 ATA 1302

RESULT 11
BC078974 1546 bp mRNA linear ROD 03-AUG-2004
LOCUS BC078974
DEFINITION Rattus norvegicus cDNA clone IMAGE:7111906, partial cde.
ACCESSION BC078974
VERSION BC078974.1 GI:50926879
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 1546)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D.,
Altchul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,D., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaefer,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McKean,P.D.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shvchenko,Y.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shvchenko,Y.,
Boutard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalske,U., Smallus,D.B.,
Schnerch,A., Schein,J.B., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
1247932
2 (bases 1 to 1546)
DIRECTOR MGC Project.
SUBMITTED (02-AUG-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.mci.nih.gov
CONTACT: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Howard Jacobs
CDNA Library Preparation: Express Genomics
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcdpax1.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Service: Ink Plate: 162 Row: F Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
FEATURES

```

source

1. .1546  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="IMAGE:7111906"  
/isue\_type="Testis, rat (Brown Norway)"  
/clone\_id="N1H\_MGC\_237"  
/lab\_host="DH10B"  
/note="Vector: pExpress1"  
144. >1546  
/codon\_start=1  
/product="Unknown (protein for IMAGE:7111906)"  
/protein\_id="AAH78974.1"  
/db\_xref="GI:50926880"  
/translation="MLSKTGAIPRPNSVGRFRRFNVOPRALFHHKLVGIET  
SCDDTAAYVDENGVGALHSQTEVHLKTVGPVPAQOOLHEHRIYBELLSAS  
GVSRSDLSAATTKPKGLALSVGLSVLVNOFKKPFPIHHEMHALITRLTHK  
VGPEFLVLSGGHCLALVQSVDFLLGKSLDIPGMDLKVARSLSLKHPECT  
MSGGKAIEHAKENRPHFTINPMONAKNDFSTFLOHTDKLITHEKEBEGIKG  
QILSSADIAAVQATACHLAKRTHRAILFCQOKNLSPANAVLVGGVANSYLIR  
RALEIVANATOCCLCPRLCTDNGIMIANGIERLBAIGLILHVDIYEPAGE  
IVSMRLVLTALTELDVSSSTHTVAHSPLNSRGANIQISMCSCTQTVNRTYRHL  
NINLEKSKKKKKKKK"

ORIGIN

Alignment Scores:

Pred. No.:	4.9e-137	Length:	1546
Score:	1725.00	Matches:	335
Percent Similarity:	91.00%	Conservative:	19
Best Local Similarity:	86.12%	Mismatches:	35
Query Match:	81.18%	Indels:	0
DB:	10	Gaps:	0

US-10-649-273-2 (1-414) x BC078974 (1-1546)

QY 1 MetleuileuthrlysthrilaglivalpnephelysProserlysarglysvaltyr 20  
DB 144 ATGCTAATGTTAAGTAAGACAGAGAGCTATTCGAGAACCAAGAGTAAGTTGCT 203  
QY 21 GlupheleuargserpnephenehiarProglythrlleupheleuhiisyllevalleu 40  
DB 204 GGAATTATAGAAATTTAATGTTCAACCTAGAGCACTTTTTCATTAATACTGGTGTG 263  
QY 41 GlylleugluphrserCyaspasptThrilaalavalasrgluthrlysalval 60  
DB 264 GGAATTAAACACAGCTGTGATGACACAGAGCTGTGTGGTGAAGAACTGGAATGG 323  
QY 61 leuuglylualalenihsersglnthrlysalvalisleysthrlygllyllevalpro 80  
DB 324 CTGGGAGAGACACGCACTCCACAGAGAGTCCATCTGAAACAGGTGGATTTCTTCT 383  
QY 81 ProAlaaglnlnleuhihsargylualansileglnargllevalglnlualseuser 100.  
DB 384 CCAAGTCCCTCAACAGCTTCACAGAGAAATATTCAGCAATATGTAAGAGAGCTTTTCT 443  
QY 101 AlhsersgylualserProserAspseuserAlaileatThrthrllyserProglyleu 120  
DB 444 GCCAGTGGGGCTCCCAAGGACCTTCAGCAATTCAGCAATCAACCAAGGAGCTG 503  
QY 121 Alaleuserleuuglylualglyleuserpnehsersleuuglnleuvalgllyleu 140  
DB 504 GCCCTAAGCCTGGGGGTTGGCTTATCCTTAGCGTACAGCTAGTAAATCAAGTTAAAAAG 563  
QY 141 Proheileprolehihsimetcylualahisaleuthrthrlleargyleuthrlysal 160  
DB 564 CCAATTATCCCAATCCATCAATGAGAGGCTCAGCGGTGACATTAATAGGCTGACCAAA 623  
QY 161 Valglupheleuvalleuvalleuvalleuvalleuvalleuvalleuvalleuval 180  
DB 624 GTGGATTTCTTTTATGATTTCTGAGGACCACTGCTGTGGCGGTAGTT 683  
QY 181 GlnlglualserAppheleuileuuglylsersleuAapilalProglysAapet 200

DB 684 CAGAGTGTTCAGATTTTCTGCTCTCGGAGAGTCCCTGGACATNAGCCGAGCGACATG 743  
QY 201 leuAaplyvalalAargyleuserleuileysheisProglylCyserThrmeser 220  
DB 744 CTGACAGGTGGAGAGAGCTTTCTTATCAACATCCGAATGTTCTCAATGAGT 803  
QY 221 Glylgllysalalileugluleuvalalysglnlyasnarphenehsapilley 240  
DB 804 GGTGGAGAAAGCTATGAACTATGGCCAAAGAGAAATAGATTCCATTTACTATCAAT 863  
QY 241 ProProleuhsheialalyeancysasppnehserspnehsersleuvallythr 260  
DB 864 CCACCACTGAGAAAGCTAGAACTGATTTTCTTTTACGGAAGCTTCAACATGTCACC 923  
QY 261 Aaplyleileileuileysglnlysgllyllylleugllysgllyllyleuser 280  
DB 924 GATTAAGCTATACACACAGAGAAAGAGAGGATTTGAGAGAGGAGGAAATCTGTCA 983  
QY 281 SerAlaAapilalAalAthrvalglnihsrmevalCyshisaleuvallysar 300  
DB 984 TCAGCGGAGACATTTGCTGCTGGTACAGACAGCAAGCGGTGACCTGGCAAAAG 1043  
QY 301 ThrhisarglaleileuupheCylysglnatgaspheleuProglinsasnalaval 320  
DB 1044 ACACATCGTCTATTTCTGTTTGGCAGAGAAATTTGCTATCTCCAGCTACGAGTA 1103  
QY 321 leuvalalaserlygllyvalalaserpnehyrthleargyAlaleuuglileu 340  
DB 1104 TTAGTGTCTCGAGAGGTTCAGAGTAACTTGACATCCAGAGACATTTGAAATTTGA 1163  
QY 341 ThransalathrlylCyserThrleuencysProProProargyleuCysthAspangly 360  
DB 1164 GCAATGCAACACATATGACTTGTGTGTCTCCCTCCGAGACTGTGCACTGCAATGG 1223  
QY 361 lleuilelealtprasnlylileglualglnalaglyleuuglylleuhsaap 380  
DB 1224 ATCATGATTCATGAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGAT 1283  
QY 381 lleugllylleargytrgluprolys 389  
DB 1284 GTAGAAACATCCGATACCAACCAAG 1310  
RESULT 12  
LOCUS AR428809 1526 bp DNA linear PART 18-DEC-2003  
DEFINITION Sequence 23 from patent US 6642041.  
ACCESSION AR428809  
VERSION AR428809.1 GI:40188595  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1526)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 23 04-NOV-2003;  
FEATURES  
source location/Qualifiers  
1. .1526  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.:	3.95e-108	Length:	1526
Score:	1385.00	Matches:	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	65.18%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2 (1-414) x AR428809 (1-1526)

QY 148 Metglualahisaleuthrthrlleargyleuthrlysalvalglupheleuval 167



|||||  
1 ATGAGGCTCAGTACCTACTACTATTAGTTGACCATATAAGATTTCTTTTATGTT 60  
168 LeuLeuLeuLeuSerGlyGlyHisCysLeuLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187  
61 CTTTGTGATTTCTGGAGGCTGCTGCTGTGTGGCATTTGACAAAGGATTCACATTTTCTG 120  
188 LeuLeuGlyLysSerSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 207  
121 CTTTCTTGGAAAGCTTTTGGACATGACACCAAGTGCATGCTTGCACAAAGTGGCAAGACA 180  
208 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGlnHis 227  
181 CTTTCTTTAATAAATCCAGAGTGTCCACCATGAGTGGGGAAGCCATAGAACAT 240  
228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247  
241 TTGGCCCAACCAAGAAATAGATTTTCATTTTGCATCAACCTCCCTTCATCATGCTAA 300  
248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267  
301 AATTGTGATTTTCTTTACTGACTGCATTCACACGTTACTGATTAATTAATGAAAAAG 360  
268 GlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 287  
361 GAAAAAGAGAGAGGATATTGAGAGAGGGCAATCTGCTTCAGACGACGACATTCCTGCC 420  
288 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 307  
421 ACAGTACAGCACACAAATGGCATGTGTCATCTGTGAAAMACACATCGGGCATTTCTGTTT 480  
308 CysLeuGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327  
481 TGTAGACAGAGAGCTTTTACTTCTCAAAATATGACAGTACTGTTGACATCTGGTGCTC 540  
328 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347  
541 GCAAGTACTCTTATATCCGACAGCTCTGGAATTTTACAAACGACACACAGTGCAT 600  
348 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 367  
601 TTGTGTGTCTCTCTCCACGACTATGCACTGATATATGCAATTATGATGCAATGCT 660  
368 IleGlnArgLeuAspAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgGlyTrp 387  
661 ATGAAAAGACTACCTGCTGCTGCTGCTGCTTTTACATGACATGAAAGGATCCGATAGAA 720  
388 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyAlaSerIleLysVal 407  
721 CCAAAATGCTCTCTTGGAGTACATATCAAAAGAGTTGAGAGAGCTTCCATTAAGTA 780  
408 ProGlnLeuLysMetGlnIle 414  
781 CCACAATTAATAATGAGAGATA 801  
RESULT 13  
BX934991 1522 bp mRNA linear VRT 02-FEB-2004  
LOCUS BX934991  
DEFINITION Gallus gallus finished cDNA, clone CHEST189114.  
ACCESSION BX934991  
VERSION BX934991.1 GI:41635519  
KEYWORDS  
ORGANISM Gallus gallus (chicken)  
SOURCE  
ORGANISM  
Gallus gallus  
Rukavortz, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 1522)  
Boardman, P.B., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.B.,  
Croning, M.D.R., Davies, R.H., Francis, M.D., Grafham, D.V.,  
Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,  
Niblett, D., Overton, I.M., Rogers, J., Scott, C.B., Taylor, R.G.,  
Tickle, C. and Wilson, S.A.

|||||  
1 ATGAGGCTCAGTACCTACTACTATTAGTTGACCATATAAGATTTCTTTTATGTT 60  
168 LeuLeuLeuLeuSerGlyGlyHisCysLeuLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187  
61 CTTTGTGATTTCTGGAGGCTGCTGCTGTGTGGCATTTGACAAAGGATTCACATTTTCTG 120  
188 LeuLeuGlyLysSerSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 207  
121 CTTTCTTGGAAAGCTTTTGGACATGACACCAAGTGCATGCTTGCACAAAGTGGCAAGACA 180  
208 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGlnHis 227  
181 CTTTCTTTAATAAATCCAGAGTGTCCACCATGAGTGGGGAAGCCATAGAACAT 240  
228 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 247  
241 TTGGCCCAACCAAGAAATAGATTTTCATTTTGCATCAACCTCCCTTCATCATGCTAA 300  
248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 267  
301 AATTGTGATTTTCTTTACTGACTGCATTCACACGTTACTGATTAATTAATGAAAAAG 360  
268 GlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 287  
361 GAAAAAGAGAGAGGATATTGAGAGAGGGCAATCTGCTTCAGACGACGACATTCCTGCC 420  
288 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 307  
421 ACAGTACAGCACACAAATGGCATGTGTCATCTGTGAAAMACACATCGGGCATTTCTGTTT 480  
308 CysLeuGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 327  
481 TGTAGACAGAGAGCTTTTACTTCTCAAAATATGACAGTACTGTTGACATCTGGTGCTC 540  
328 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347  
541 GCAAGTACTCTTATATCCGACAGCTCTGGAATTTTACAAACGACACACAGTGCAT 600  
348 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 367  
601 TTGTGTGTCTCTCTCCACGACTATGCACTGATATATGCAATTATGATGCAATGCT 660  
368 IleGlnArgLeuAspAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgGlyTrp 387  
661 ATGAAAAGACTACCTGCTGCTGCTGCTGCTTTTACATGACATGAAAGGATCCGATAGAA 720  
388 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyAlaSerIleLysVal 407  
721 CCAAAATGCTCTCTTGGAGTACATATCAAAAGAGTTGAGAGAGCTTCCATTAAGTA 780  
408 ProGlnLeuLysMetGlnIle 414  
781 CCACAATTAATAATGAGAGATA 801  
TITLE  
JOURNAL  
COMMENT  
Direct Submission  
Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: chickens@ms.umbist.ac.uk  
BBSRC/Bundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA  
sequencing project.  
This sequence is from the  
BBSRC/Bundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
from a library constructed by Elizabeth Bosch. cDNA was prepared  
from RNA extracted from muscle, normalized, and poly A-tailed.  
EcoRI-NciI cut cDNA was then ligated into the vector. Vector:  
pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NciI Host: Escherichia  
coli DH10B.  
FEATURES  
source  
1..1522  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Layer and broiler"  
/db\_xref="taxon:9031"  
/clone="CHEST189114"  
/clone\_1ib="CSFORBN11"  
/dev\_stage="adult"  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,56e-106 Length: 1522  
Score: 1362.00 Matches: 258  
Percent Similarity: 83.20% Conservative: 54  
Best Local Similarity: 68.80% Mismatches: 63  
Query Match: 64.09% Indels: 0  
Gaps: 0  
US-10-649-273-2 (1-414) x BX934991 (1-1522)  
QY 37 LysIleValLeuGlyIleGlnThrSerCysAspAspThrAlaAlaValValAspGlu 56  
DB 254 AACCTTGCTGCTGGGATCAACCAACGCTGTGACACACGGCGCGGCTGTGACGAG 313  
QY 57 ThrGlyAsnValLeuGlyGlnAlaIleHisSerGlnThrGlnValHisLysLysThrGly 76  
DB 314 GCGGCGACGCTGCTGGAGAACGCTGACAGCCAAAGAGGTCCACTCAAGCAGGT 373  
QY 77 GlyIleValProProAlaAlaGlnGlnLeuHisArgGlnAsnIleGlnArgIleValGln 96  
DB 374 GGAATTAATTCCTCAGTACGACACGACGCTTCACAGAGAAAGCATCCAGCAAGTAAAG 433  
QY 97 GlnAlaLeuSerAlaSerGlyValSerProSerAspLysSerAlaIleAlaThrTrpIle 116  
DB 434 GAAGCAGTACAGTCCAGTGGAGTTCCTGTAATGAACCTGCTGCTATTCACATACAGTG 493  
QY 117 LysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGly 136  
DB 494 AAACCGAGACTTGCCTGACGCTGAGGTGGAGCTGCACATGAGCTTACAGCTGGAGTAC 553  
QY 137 GlnLeuLysLysProPheIleProIleHisHisMetGlnAlaHisAlaLeuThrIleArg 156  
DB 554 AGGTACCGAAGCCTTTCATACCATTCATCAATGAGAGCTCAGCAGCTTACATACAGA 613  
QY 157 LeuThrAsnLysValGlnPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeu 176  
DB 614 CTGACAGAGCAAGTAAATTTCTCTTCTTAACTCTCCGAGAGTCACTGCATC 673  
QY 177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 196  
DB 674 TTGGCAGTAGACGACGAGGATTCACATTTCTTCTGCTTGGACAGTCCATAGATATAGCA 733  
QY 197 ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCys 216  
DB 734 CCAAGGTGACATGCTGATTAAGTACGAAAGAGCTCTTTAGTGAAGACCCGAGTGC 793  
QY 217 SerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnIleLysAsnArgPheHis 236  
DB 794 CACGCGATGCTGCGGGGAGAGGCAATGAGCACTGCTCAACCGAGACTGCGCAACG 853



QY	Phasapri1elyrProPtolelehihi1salalyasanCySaAaPhPeSePheThGlYeu	256
QY	237 Pheapri1elyrProPtolelehihi1salalyasanCySaAaPhPeSePheThGlYeu	256
Db	854 TACACTTCAGACTTCCTCCATGACACAGATACGATACCTGATATTTCTTCTCCGACTT	913
QY	257 Glnhi1sValThrAspLy1leilemetysylsglysgluGluGly1leGluysglY	276
Db	914 CAGAGCCTTCACACAAAGCCATTCTTCAGAAAGAAAAGAAAGAAAGATTCACAGAACGG	973
QY	277 Gln1leleuSerSer1alalalasp1lealalathrValGlnhi1sThrMetAlaCysHi	296
Db	974 GAAATCCCTGCTCTGCGTTAAGGACATCCCTGCTGCTGCACAGACAGTACGTAGCGTCCAT	1033
QY	297 LeuVallysaAggThrhi1sArgAlalaleuPheCyslysglnArgAspLeuLeuProGln	316
Db	1034 ATTTATCCAGCGACACACCCAGGACCGATCGCTTCTTCATGAAAACACAGCATTTATTACCA	1093
QY	317 AsnAsnAlaValleuValAlaserGlyValAlaserAsnPhenyl1argArgAla	336
Db	1094 AAAAGTCAACTCTGCTGTTCTATCAGAGAGAGTTCGAATTAATCAGTATTCAGAAAAGA	1153
QY	337 LeuGlu1leuThrAsnAlaThrglnCysThrleuLeuCysProProArglyCys	356
Db	1154 CTGCAAGACTCTGCAGAAATGCAAAACGGTTTGGCTTTCTGTCCTCCCTCCAGGCTGTGC	1213
QY	357 ThrThrapengly1leMet1lealathrapengly1leGluargLeuArgAlaGlyleuGly	376
Db	1214 ACCGATATATGTGTATTATATTCATGACAGATGCATTTGAAGGTTCGCGACGAGATGTGCT	1273
QY	377 Ileleuhi1sAsp1leGluGly1leargYtrGluPPolysCysProLeuGlyValAsp1le	396
Db	1274 ATTTTATTCAGTACTGATCGATCGATCCGCTATCCAGAACCAAAAGCTCCCTTCGATATGATTT	1333
QY	397 SerlysgluVal1glyGluAlaser1lelyValProGlnleuLeu 411	
Db	1334 TCCAAAGAGTTGAAGAGATTCATCAAGAGTGAAGCCAGACATTAGG 1378	
RESULT 14		
LOCUS	BC038910	
DEFINITION	Mus musculus O-sialylcoprotein endopeptidase-like 1, mRNA (cDNA	
ACCESSION	BC038910	
VERSION	BC038910.1	
KEYWORDS	GI:24433548	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1017)	
AUTHORS	Straubenberg,R.L., Fellings,D.E.A., Grouse,L.H., Dergs,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Butcov,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Riddin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Schaefer,T.M., Brownstein,M.J., Udell,T.B., Toshitsuyuki,S., Carinci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Wotley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hilyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,B., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouford,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,J.S., Krzyzaniak,M.I., Skalska,U., Smalins,D.E., Schermer,A., Schein,J.B., Jones,S.J. and Marra,M.A.	
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
MEDLINE	22388257	
PUBMED	12477932	
REFERENCE	2 (bases 1 to 1017)	

**AUTHORS** Strausberg, R.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (25-Oct-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK** NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
**COMMENT** Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada  
[info@cgsc.bc.ca](mailto:info@cgsc.bc.ca)  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letitia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice Mcleavy, Steven Nees, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stolt, Michael Thorne, Miranada Tsai, Natalya van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

**FEATURES**  
 source  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAX Plate: 86 Row: C Column: 12.  
 Location/Qualifiers  
 1..1017  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5053559"  
 /tissue\_type="Liver, normal, 5 month old male mouse."  
 /clone\_lib="NCI CGAP\_L19"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

**ORIGIN**  
 Alignment Scores:  
 Pred. No.: 7,91e-99 Length: 1017  
 Score: 1273.00 Matches: 241  
 Percent Similarity: 92.73% Conservative: 14  
 Best Local Similarity: 87.64% Mismatches: 20  
 Query Match: 59.91% Indels: 0  
 Db: 10 Gaps: 0

US-10-649-273-2 (1-414) x BC038910 (1-1017)

QY	140	LYsPProPheIlePProIleHsHsMetGluAlaHsAlaLeuThrIleArgLeuThraen	159
Db	2	AAAGCATTGATCCCGATTCATCATGAGGTCAAGCATGACATATTAAGGTCAACAT	61
QY	160	LYsValGluPhePProPheLeuValLeuLeuIleSerGlyIYHsCysLeuLeuAlaLeu	179
Db	62	AAAGTAGAATTTCTTTTATGTTCTTTGATTTCTGGGGCTCACTGCTGTGGCATTA	121
QY	180	ValGlnGlyValSerAspPheLeuLeuLeuGlyLYsSerLeuAspIleAlaPProGlyAsp	199
Db	122	GTCAGAGGTGTTCCGATTTCTGCTCTTGGAAGTCTTGGACATGACATGACACGCGAC	181
QY	200	MetLeuAspLYsValAlaArgArgPheSerLeuIleLYsHsIlePProGluCysSerThrMet	219
Db	182	ATGCTTGACAGGTGGCGCAAGAAGCTTCTTAAATCAACATCCAGAAAGTCTTCAAGT	241
QY	220	SerGlyGlyLYsAlaIleGluHsLeuAlaLYsGlnGlyAsnArgPheHsIlePheAspIle	239
Db	242	AGGTGTGGAAGAGCTATGAAACAGTTGGCCAAAGACGAAATAGATTCCATTTTACTATC	301
QY	240	LYsPProPProLeuHsHsAlaLYsAsnCYsAspPheSerPheThrGlyLeuGlnHsVal	259
Db	302	AATCCACCTAAGCAGAAAGCTTAAGAAATGCGATTTTCTTACGCGACCTTCAACATATT	361



Db 1386 GATGTGTATTATTATACAGTACTGATGCGCATCGCTATGAACCAAGCTCCCCCTTGA 1445  
Qy 394 ValAspIleSerIysGluValGlyGluAlaSerIleIysValProGlnLeuIys 411  
Db 1446 ATTGATATTTCCAAGAGGTGAAGAGGATTCCATCAGATGCGCAAGACTAAGG 1499

Search completed: February 16, 2005, 18:02:36  
Job time : 6632 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:09:32 ; Search time 788.4 Seconds  
(without alignments)  
3108.540 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 2125  
Sequence: 1 MLILTKAGVFFKPSKRVY.....DISKEVBSIKVPLKMEI 414

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4390206 segs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seg length: 0  
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=x1h  
-Q=cg21\_1/USPTO.spool/US10649273/runat\_14022005\_114702\_16389/app\_query.fasta\_1.1429  
-DB=N Geneseq\_16Dec04 -OPMT=fastcap -SUFFIX=ing -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US10649273.@CGN\_1\_11057.@runat\_14022005\_114702\_16389 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=0.5 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_16Dec04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	2058	6	ABA93268 Human O-s
2	2125	100.0	2197	6	AB276635 DNA encod
3	2125	100.0	2572	8	ABT23207 Human pro
4	2088	98.3	1820	6	AA46856 Human g1y
5	2088	98.3	1820	8	ACA60887 Human CDN

6	2088	98.3	1821	10	AB557020	Ab557020 cDNA enco
7	1944	91.5	2208	10	ADA52832	Ada52832 Human cod
8	1944	91.5	2890	12	ADQ24627	Adq24627 Human sof
9	1747	82.2	1416	8	ABX70950	Abx70950 Novel hum
10	1549	72.9	1572	6	ABQ75508	Abq75508 Murine si
11	1385	65.2	1526	6	ABQ75639	Abq75639 DNA encod
12	1001.5	47.1	1338	10	ADB31345	Adb31345 Human d1a
13	714.5	33.6	1601	4	AB124633	Ab124633 Drosophi
14	700.5	33.0	1385	4	AAH15110	Aah15110 Human CDN
15	687.5	31.2	3656	4	AB124632	Ab124632 Drosophi
16	662.5	31.2	1557	3	ABC38454	Abc38454 Arabidops
17	629	29.6	571	4	AAH12983	Aah12983 Human CDN
18	599	28.2	2734	5	AA584622	AA584622 DNA encod
19	531.5	25.0	4360	6	AA48239	Aad48239 Ehrlichia
20	524	24.7	1146	8	ACA26804	Aca26804 Prokaryot
21	502	23.6	1044	8	ACA39102	Aca39102 Prokaryot
22	502	23.6	94750	4	AA28551	AA28551 Genomic f
23	501	23.6	1053	12	ADL03120	Adl03120 DNA encod
24	500	23.5	1092	6	ABQ90383	Abq90383 M. capsul
25	499.5	23.5	1206	11	ABD02197	Abd02197 Pseudomon
26	494.5	23.3	1032	8	ACA43173	Aca43173 Prokaryot
27	492.5	23.2	1026	4	AA554064	AA554064 Pseudomon
28	492.5	23.2	1026	8	ACA42146	Aca42146 Prokaryot
29	492.5	23.2	1026	10	ADG73341	Adg73341 P aerugin
30	492.5	23.2	1059	11	ABD02280	Abd02280 Pseudomon
31	490.5	23.1	1026	10	ADG73343	Adg73343 P aerugin
32	484.5	22.8	1014	4	AA552570	AA552570 E. coli D
33	484.5	22.8	1014	8	ACA51431	Aca51431 Prokaryot
34	484.5	22.8	1014	8	ACA32641	Aca32641 Prokaryot
35	483.5	22.8	1014	8	ACA54107	Aca54107 Prokaryot
36	482.5	22.7	1020	8	ACA44384	Aca44384 Prokaryot
37	482.5	22.7	1074	10	ADR02056	Adr02056 Bacterial
38	481.5	22.7	1014	4	AA556045	AA556045 Salmonell
39	479	22.5	1315	2	AAQ27645	Aaq27645 Glycero
40	475.5	22.4	1029	4	AA553309	AA553309 Haemophil
41	475.5	22.4	1029	8	ACA34150	Aca34150 Prokaryot
42	475.5	22.4	110000	2	AA742063	AA742063 Continuat
43	474.5	22.3	1032	11	AC96525	Ac96525 Klebsiell
44	472.5	22.2	1014	10	ACF71364	Acf71364 Phototrab
45	472.5	22.2	110000	10	ACF67367_45	Continuation (46 o

## ALIGNMENTS

RESULT 1	ABA93268	standard; cDNA; 2058 BP.
ID	ABA93268	
XX	ABA93268;	
AC		
XX		
DT	19-APR-2002	(first entry)
XX		
DE	Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.	
XX		
KM	Human; O-sialoglycoproteinase-like protein; OSGPAP; enzyme; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	110..1354
FT		/tag= "a
FT		/product= "O-sialoglycoproteinase-like protein"
XX		
PN	CN1318550-A.	
PD		
XX	24-OCT-2001.	
XX		
PF	19-APR-2000; 2000CN-00106834.	
XX		
PR	19-APR-2000; 2000CN-00106834.	
XX		
PA	(SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.	
XX		

PI Mao Y, Xie Y;  
 XX WPI: 2002-115090/16.  
 DR P-PSDB; ABB05481.  
 XX  
 PT O-sialosylcoproteinase-like protein and encoding polynucleotide, useful  
 for diagnosing, preventing and treating related diseases.  
 XX  
 PS Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.  
 XX  
 CC The present sequence encodes human O-sialosylcoproteinase-like protein  
 CC (OSGRLP). The present invention also describes: (1) the preparation of  
 CC the OSGRLP protein; (2) applying the OSGRLP protein in diagnosis; (3) the  
 CC prevention and/or treatment of related diseases; (4) utilizing the OSGRLP  
 CC protein in screening its agonist, excitomotor and inhibitor and preparing  
 CC an antibody against the OSGRLP protein; and (5) the use of the OSGRLP  
 CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors  
 CC and antibodies in treating diseases related to the abnormal OSGRLP gene  
 CC and in preparing the medicine composite for the treatment  
 CC  
 SQ Sequence 2058 BP; 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 1,946-212 Length: 2058  
 Score: 2125.00 Matches: 414  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-10-649-273-2 (1-414) x ABA93268 (1-2058)  
 QY 1 MetLeuIleuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
 DB 110 ATGCTAATCTTGACTTAAGACCTGCAGAGATTTTAAACCATCAAAAGAAAGTTTAT 169  
 QY 21 GluPheLeuArgSerPheAsnPhelAspProGlyThrLeuPheLeuHisLysIleValLeu 40  
 DB 170 GAATTTTAAAGAGTTTAAATTTTCTCTGGAACACATTTCTTCATTAATATGATTG 229  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60  
 DB 230 GGAATGAACTAGTGTGATGATACAGACAGCTGCTGTGTGATGAAGAATCTGAATATG 289  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIleValPro 80  
 DB 290 TTGGAGAGAGCAATACATTTCCAACTGAAGTTCAATTTAAACAGGTGGATTTGCTCT 349  
 QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnAspIleValGlnGluAlaLeuSer 100  
 DB 350 CCAGACAGCTCAACAGCTTCAACAGAGAAATATTCAACGATATGTCACAGAGCTCTTTCT 409  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
 DB 410 GCCAGTGAAGCTCTCCCAAGTGACCTTCAGCAATTTGCAATACATAAACAGGACTT 469  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140  
 DB 470 GCTTTAAGCCCTGGAGGTGGCTTTATCATTTTACCTTACAGCTGTGTGAGCAAGTTAAAG 529  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
 DB 530 CCAATTCATCCCATTCATCATATGAGGCTCATGCACTTACATTTAGGTGACCAATAAA 589  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 DB 590 GTAGAATTTCTTTTATTTAGTTCTTTGATTTCTGAGAGCTCATGCTGTGTGCAATTAGTT 649  
 QY 181 GlnGlyValSerAspPheLeuLeuGlnGlyLysSerLeuAspIleAlaProGlyLysMet 200  
 DB 650 CAAGAGATTTTCAAGATTTTCTGCTCTTGAAAGTCTTTTGACATAGCACAGGTGACATG 709  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluLysSerThrMetSer 220

DB 710 CTTCACAGGTGCAAGAGACTTCTTAAATCAATCCACAGTGTCTCACCATAGT 769  
 QY 221 GlyGlyLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 DB 770 GGTGGGAAAGCCATTAACATTTGGCCAAACAGAAATATGATTTTACATTTGACATCAAA 829  
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 DB 830 CTTCCCTTCATCATCATCTAAATATGATTTTCTTTTACTGAGCTTCAACAGCTTACT 889  
 QY 261 AspLysIleIleMetLysLysGlyLysGlyGlyIleGlyLysGlyGlnIleLeuSer 280  
 DB 890 GATTAATATATATGAAAAAGGAAAAAGGAAAGGATTTGACAGAGGGCAAACTCTGCT 949  
 QY 281 SerAlaAlaAspIleAlaThrValGlnHisIleThrMetAlaCysHisLeuValLysArg 300  
 DB 950 TCAGCAGCAGACATTTCTCCACAGTACAGACACACAAATGCAATGCTATCTTGTGAAAGA 1009  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
 DB 1010 ACACATCGGGCTATTCTGTTTGTAAACAGAGACTTGTACTCAAAATTAATGCACTA 1069  
 QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPhelTyrIleArgArgAlaLeuGluIleLeu 340  
 DB 1070 CTGGTTGCACTGTGTGTGTGCGCAAGTAATCTTATATCCGACAGCTCTGAAATTTTA 1129  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgArgLeuCysThrAspAsnGly 360  
 DB 1130 ACMAAGCAACACAGACACTTGTGTGTCTCCCTCCACAGCTATGACCTGAATAAAGGC 1189  
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1190 ATTATATATTCATGAGAAATGATTTGAAGACTACGCTGCTGGCATTTTATCATGAC 1249  
 QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGlyVal 400  
 DB 1250 ATAGAGGCAATCGCATAGAACCAAAATGTCCTTGAGATAGCATATCAAAAGAGTT 1309  
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 DB 1310 GGAAGAGCTTCCATTAAGTACCAACAATTAATAATGAGATA 1351  
 RESULT 2  
 ABS76635  
 ID ABS76635 standard; DNA; 2197 BP.  
 XX  
 AC ABS76635;  
 XX  
 DT 11-DEC-2002 (first entry)  
 XX  
 DE DNA encoding novel human metalloprotease MPI.  
 XX  
 XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 KW neurological disorder; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200272751-A2.  
 XX  
 XX 19-SEP-2002.  
 PD  
 XX 05-FEB-2002; 2002WO-US003353.  
 PF  
 XX 05-FEB-2001; 2001US-0266518P.  
 PR

PR 10-APR-2001; 2001US-0282814P.  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX PI Chen J, Feder J, Nelson TC, Duclos F, Kyteek S;  
 XX WPI; 2002-723329/78.  
 DR P-PSDB; ABG96478.  
 XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
 PT treating, or ameliorating diseases associated with aberrant  
 PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
 PT neurological disorders.  
 XX  
 XX Claim 1, Fig 1A-C; 473pp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC metalloproteinase (MP-1). (I) is useful for preventing, treating, or  
 CC ameliorating a medical condition, particularly an immune disorder, an  
 CC aberrant glutamate transport or motor neuron disorder, such as  
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
 CC condition. The compositions and methods are also useful for diagnosing,  
 CC prognosticating, treating, ameliorating and/or treating disorders  
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease). This sequence represents a  
 CC metalloproteinase MP1 polynucleotide  
 XX  
 SQ Sequence: 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 2,14e-212 Length: 2197  
 Score: 2125.00 Matches: 414  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-10-649-273-2 (1-414) x ABS76635 (1-2197)  
 QY 1 MetLeuLeuLeuThrLysThraGlyValPhePheLysProSerLysArgLysValTyr 20  
 DB 231 ATGCTAATCTTGAAGACTGAGAGAGTTTATTTTAAACATCAAAAAGAAAGTTTAT 290  
 QY 21 GluPheLeuArgSerPheLeuPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
 DB 291 GAATTTTAAAGAGTTTAAATTTTATCTCTGGAACACATTTCTTCATTAATAATAGTATG 350  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaIleValIleAspGluThrGlyLysVal 60  
 DB 351 GGAATTAAGAACTAGTGTGATGATACAGACAGCTGTGTGGTGAATGAACCTGAAATG 410  
 QY 61 LeuGlyValAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyValPro 80  
 DB 411 TTGGGAGAGCAATACATTTCCCAACGAGTTCATTTTAAACAGGCGGAGTTGTTCT 470  
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 471 CCAGCAGCTCAACAGCTTCACAGAGAAAATATTCACAAATAGTACAGAGAGCTTTCT 530  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
 DB 531 GCCAGTGAAGTCTCTCCAAAGTGAACCTTCAGCAATTCGACATTAACCAAGAGACTT 590  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysVal 140  
 DB 591 GCTTTAAGCTTGGAGTGGAGCTTATCATTTTACGTTACAGCTGGTGAACAGTTAAAGAG 650

QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAspLys 160  
 DB 651 CCATTCATTCCTCCATTCATCATATGAGAGCTCATGCACTTACTATTAAGTTACCAATTA 710  
 QY 161 ValGluPheProPheLeuValIleLeuLysSerGlyValHisCysLeuLeuAlaLeuVal 180  
 DB 711 GTGAATTTCTCTTTTATGTTCTTTTGAATTTCTGAGAGTCACTGTCTGGGCAATTAGT 770  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyLysMet 200  
 DB 771 CAAGAGTTTCAATTTTCTGCTTTTGAAGAGCTTTTGAACATGACCAAGGAGCATG 830  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
 DB 831 CTTCAGAGGTGCAAGAGAGCTTTCTTTATTAATCAATCAGAGTCTCCACCATAGT 890  
 QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 DB 891 GGTGGAGAAAGCCATGAGACATTTGGCCAAACAGAAATAGATTTCATTTGACATCAA 950  
 QY 241 ProProLeuHisHisAlaLysValCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 DB 951 CCTCCCTTCATCATGCTAATAATTCGATTTTCTTTTCTGAGCTTCAACAGTTACT 1010  
 QY 261 AspLysIleIleMetLysLysGluLysGluGlnGlyIleGlnLysGlyGlnIleLeuSer 280  
 DB 1011 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1070  
 QY 281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
 DB 1071 TACGACGACAGACTTCTCTGCAAGTACAGACACATGCAATGCAATGCTTGTGAAGA 1130  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
 DB 1131 ACACATCGGGCTATTCGTTTGTGAAGCAGAGACTTGTACTCAAAATTAATGACGTA 1190  
 QY 321 LeuValAlaSerGlyValValAlaSerAsnPheThrIleArgArgAlaLeuGluIleLeu 340  
 DB 1191 CTGGTTCATCTGGTGTGTGTGCGAAGTAACTTTATATTCGACAGAGCTCTGGAATTTTA 1250  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360  
 DB 1251 ACMAAGCAACACAGTGCATTTGTTGTCTCTCTCCAGACTATGACATGTAATGCG 1310  
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1311 ATTATGATTGCAATGGAATGATTAAGAACTACGTCGTGGCTTGGCAATTTTACATGAC 1370  
 QY 381 IleGluGlyIleArgGlyGluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
 DB 1371 ATAGAAGGATCGCATTAACCAAAATATGCTCTTGGAGTACATATCAAAAGAAATT 1430  
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGlnIle 414  
 DB 1431 GGAAGAGCTTCCATTAATAAGTACCAATTAATAATGAAGATA 1472

RESULT 3  
 ABT23207  
 ID ABT23207 standard; DNA; 2572 BP.  
 XX  
 XX ABT23207;  
 AC  
 AC 01-MAY-2003 (first entry)  
 DT  
 DT Human protein modification + maintenance molecule DNA SEQ ID No 36.  
 XX  
 XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
 KW cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
 KW anticancer; hepatoprotective; gynaecological; antibacterial; virucide;  
 KW protozoacide; antiparasitic; cell proliferative diseases; PMOD;  
 KW protein modification and maintenance molecule; immunogenic fragment;  
 KW cancer; autoimmune; inflammatory disease; neurological disorder;

KM gastrointestinal; developmental; vesicle trafficking disorder; infection;  
KM protein-protein interaction; drug-target interaction;  
KM gene expression profile; human; gene; de.  
XX Homo sapiens.  
PN MO200300844-A2.  
PD 03-JAN-2003.  
PE 18-JUN-2002; 2002WO-US019360.  
PR 22-JUN-2001; 2001US-0300508P.  
PR 06-JUL-2001; 2001US-0303445P.  
PR 13-JUL-2001; 2001US-0305405P.  
PR 09-AUG-2001; 2001US-0311442P.  
PR 24-AUG-2001; 2001US-0314821P.  
PR 29-AUG-2001; 2001US-0315992P.  
PR 03-MAY-2002; 2002US-0378205P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
PI Gandhi AR, Kabie AE, Swarnakar A, Hafalia AA, Tran B, Duggan BM,  
PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DAM, Lee EA, Yue H,  
PI Forsythe J, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J,  
PI Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Borowsky ML, Yao MG,  
PI Walia NK, Mason PM, Gururajan R, Lee S, Becha SD, Lee SY, Tran UK,  
PI Elliott VS, Luo W, Sprague WM, Tang YT, Lu Y, Zebardian Y;  
XX  
DR WPI: 2003-184039/18.  
DR P-PSDB; ABJ26554.  
PT New isolated human PMOD polypeptide and polynucleotide, useful for  
PT diagnosing, treating and preventing diseases or conditions associated  
PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  
PT infections.  
XX  
PS Claim 91; Page 211; 225pp; English.  
XX  
CC The invention relates to an isolated polypeptide comprising: any of 28  
CC sequences of 48-1256 amino acids; a natural amino acid sequence at least  
CC 90% identical to the 28 amino acid sequences, 94% identical to a sequence  
CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
CC acids, or 97% identical to a sequence of 242 amino acids, all given in  
CC the specification; or a biologically active or immunogenic fragment of  
CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
CC in diagnosing, treating and preventing diseases or conditions associated  
CC with the decreased expression of protein modification and maintenance  
CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
CC endometriosis), developmental, vesicle trafficking disorders, and  
CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
CC useful in assessing the effects of exogenous compounds on the expression  
CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
CC fragments are useful in screening compounds for effectiveness as agonist  
CC or antagonist of the polypeptides, or in altering the expression of the  
CC target polynucleotide and compounds that specifically bind to or modulate  
CC the activity of the polypeptide. The microarray is useful in monitoring  
CC or measuring protein-protein interactions, drug-target interactions, and  
CC gene expression profiles. This polynucleotide sequence represents the DNA  
CC encoding a human PMOD protein of the invention  
XX  
SQ Sequence 2572 BP; 780 A; 489 C; 509 G; 794 T; 0 U; 0 Other;  
XX  
Alignment Scores:  
Pred. No.: 2.7e-212 Length: 2572  
Score: 2125.00 Matches: 414  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-10-649-273-2 (1-414) x ABR23207 (1-2572)  
QY  
1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLysProSerIleArgLysValIleTyr 20  
DB  
144 ANGCTAATCTGACTAAGACTGACGAGATTTTTTTTAAACCAACAAAGGAAGTTTAT 203  
QY  
21 GluPheLeuArgSerPheAsnProHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
DB  
204 GAATTTTAAAGAGTTTAAATTTTCACTCGAAGCACTATTTCTTCAATAAATAGATTG 263  
QY  
41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60  
DB  
264 GGAAATTGAAGAACTAGTTGTGATGATACAGCAGCTGCTGTGTGATGAATACTGAAATGTG 323  
QY  
61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIleValPro 80  
DB  
324 TTGGGAGAAACAAATACATTCGCCAAGCTGAAGTTCAATTAAACAGGTGGATTTGCTT 383  
QY  
81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
DB  
384 CAGAGAGCTCAACAGCTTCAAGAGAAATATTCAAGAAATGTAACAAGAGCTCTTTCT 443  
QY  
101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
DB  
444 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATTGCAATCAATCAATCAATCAATCAAT 503  
QY  
121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyLeuLysLys 140  
DB  
504 GCTTTAAGCTCGAGAGTGGGCTTATCATTTAGCTTACAGCTGTAGACAGTTAAAG 563  
QY  
141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
DB  
564 CCAATTATTCCTCAATTCATATGAGAGCTCAAGCACTTATATGAGTGAACCAATAAA 623  
QY  
161 ValGluPheProPheLeuValLeuLeuIleSerGlyValHisCysLeuLeuAlaLeuVal 180  
DB  
624 GTRGAATTCCTTTTATGTTAGTCTTTGATTTCGAGAGTCACGTCTGTTGGCATTAAGT 683  
QY  
181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
DB  
684 CAAGAGTTTCAGATTTTCTGCTTGGAAAGTCTTTGGACATPACACACAGGTGACATG 743  
QY  
201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
DB  
744 CTTGACAAAGTGCAGAAAGACCTTTTAATAAATCCCAAGTGCCTCCACCAATGAGT 803  
QY  
221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
DB  
804 GGTGGAAAGCCATTAACATTTGGCCAAACAGAAATAGATTTTCAATTTGACATCAAA 863  
QY  
241 ProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeuGlnHisValThr 260  
DB  
864 CCGCCCTTGATCATGCTAAATATGATTTTCTTTTACGTGACCTCAACAGTTACT 923  
QY  
261 AspLysIleIleMetLysLysGlyLysGluGluGlyIleGluLysGlyGlnIleLeuSer 280  
DB  
924 GATTAATATATATGAAAAAGAAAAAGGAAGGATTTTCAAGAAAGGGCAAACTCTCTCT 983  
QY  
281 SerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
DB  
984 TCAAGCAGACATCTCTGACAGATGACAGACATGCAATGCAATGCTTTGTGAAAGAA 1043  
QY  
301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
DB  
1044 ACAATCTGGGCTATTTCTGTTTCTTAAGCAGAGAGACTGTATTCTCAAAATTAATGCGATA 1103  
QY  
321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeu 340  
DB  
1104 CTGGTTGCATCTGGTGTGTGTGCGAAGTAATCTTATATCCGACAGAGCTCGGAAATTTTA 1163  
QY  
341 ThrAsnAlaThrGlnCysThrIleLeuLysProProProArgLeuCysThrAspAsnGly 360





QY 241 ProProLeuHisHisAlaIyAsnCyAspPheSerPheThrGlyLeuGlnHisValThr 260  
 DB 866 CTTCCCTTCATCTATCTAATAATTTGATTTTCTTTTACGTGACCTTCAACACCTTACT 925  
 QY 261 AspIleValIleMetLeuValGluValGluGluGluGluGluGluGluGluGluGluGlu 280  
 DB 926 GATTAATAATATATGAAAAACAGAAACAGAGAGAGATTTGAGAGAGAGAGAGAGAGAGAG 985  
 QY 281 SerAlaIleAspIleAlaIleThrValGlnHisSerThrMetAlaGlyHisLeuValIleVal 300  
 DB 986 TCAGACACAGACATCTGCTGACAGTACAGACACAAAGGACATCTCTCTGTAAGAAGA 1045  
 QY 301 ThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
 DB 1046 ACACATGGGGCTATTTCTTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1105  
 QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeu 340  
 DB 1106 CTGGTTCATCTGGTGGTGGTGGCAAGTAACTTCTATATCCGACAGAGCTCTGAGAAATTTTA 1165  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuGlyProProProProArgLeuGlyThrAspAsnGly 360  
 DB 1166 ACAAAACGACACACAGTGCATCTTGTGTCTCTCTCCACAGACTATGACATGATATATGAC 1225  
 QY 361 IleMetIleAlaIleTPaenGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1226 ATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1285  
 QY 381 IleGluGlyIleAlaGlyTyrGluProGlyCysProLeuGlyValAspIleSerIleGluVal 400  
 DB 1286 ATGAAAGGCAATCCGCTATGAAACCAAAAGTCTCTTGGAGTGAATATCAAAAGAGATT 1345  
 QY 401 GlyGluAlaSerIleLeuValProGlnLeuLeuMetGluIle 414  
 DB 1346 GGAAGAGCTTCATTAATAAGTACCAATTAATAAGTACGAGATA 1387

## RESULT 5

ACAG0887 ID ACAG0887 standard; cDNA; 1820 BP.

ACAG0887; XX

08-JUL-2003 (first entry)

Human cDNA 28472 encoding a glycoprotease.

Human; sex; gene; cancer; aberrant cellular proliferation;  
 differentiation; immune disorders; heart disorder; brain disorder;  
 cardiovascular disorder; endothelial cell disorder; pain disorder;  
 haematopoietic disorder; blood vessel disorder; metabolic disorder;  
 liver disorder; platelet disorder; glycoprotease.

Homo sapiens.

Key Location/Qualifiers

FT 146..1390

FT /product= "Glycoprotease"

FT /note= "This CDS is specifically claimed in claim 1"

PN US2003009017-A1.

XX 09-JAN-2003.

XX 08-NOV-2001; 2001US-00012140.

XX 08-NOV-2000; 2000US-0246768P.

XX 08-NOV-2000; 2000US-0246772P.

XX 15-NOV-2000; 2000US-0249185P.

XX (LEIB/) LEIBY K R.

XX (KAPF/) KAPFELER-LIBERMANN R.

PA (GLUC/) GLUCKSMANN M A.  
 XX Leiby KR, Kapellier-Libermann R, Glucksmann MA;  
 PI WPI; 2003-428888/40.  
 XX P-PSDB; AB009569.  
 DR New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
 XX molecules, useful for diagnosing, treating cancer, pain, or immune,  
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic  
 PT and liver disorders.  
 PS Claim 2; Fig 8; 90P; English.

The invention relates to an isolated 38650 (encoding adenosine  
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
 CC a sequence which is at least 60% identical to the six nucleic acids or  
 CC their open reading frames, fragments of at least 15 nucleotides,  
 CC naturally occurring variants, or a DNA insert of the plasmid deposited  
 CC with the American Type Culture Collection as Accession No. not defined in  
 CC the specification, which encodes the amino acid sequence). Also included in  
 CC are a host cell containing the nucleic acids (used to produce the  
 CC proteins), the encoded proteins, an antibody that selectively binds to  
 CC the polypeptide, and identifying a compound that binds to/modulates the  
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
 CC methods are useful for diagnosing, treating cancer, aberrant cellular  
 CC proliferation and/or differentiation, immune disorders, heart disorders,  
 CC cardiovascular disorders including endothelial cell disorders,  
 CC haematopoietic disorders, blood vessel disorders, brain disorders, pain  
 CC and metabolic disorders, liver disorders and platelet disorders (many  
 CC examples of these disorders are given in the specification). The present  
 CC sequence is the Human cDNA 28472 encoding a glycoprotease

XX Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;

## Alignment Scores:

pred. No.: 1 246-208 Length: 1820  
 Score: 2088.00 Matches: 407  
 Percent Similarity: 99.03% Conservative: 3  
 Best Local Similarity: 98.31% Mismatches: 4  
 Query Match: 98.26% Indels: 0  
 DB: Gaps: 0

US-10-649-273-2 (1-414) x ACAG0887 (1-1820)

QY 1 MetLeuIleLeuThrHisThrAlaGlyValPhePheLeuProSerIleArgValIleTyr 20  
 DB 146 ATGCTAATCTTGACTAAGACTGACGAGAGTTTATTAACCATCAAAAGAAAGTTTAT 205  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValIleLeu 40  
 DB 206 GAATTTTAAAGAGTTTAAATTTTCACTCGGAACCTATTTCTCTCAATAAATGATTTG 265  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaValAspGluThrGlyAsnVal 60  
 DB 266 GGAATTAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 325  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisIleLeuValSerThrGlyGlyIleValPro 80  
 DB 326 TTGGAGAGAGCATATCTCCCAAGCTGAAGTTCAATTAATAAAGAGTGGAGATTGTTCT 385  
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAlaLeuSer 100  
 DB 386 CCAGACAGCTCAACAGCTTCAACAGAAATATTCACGAATAGTACAAAGAGCTCTTCT 445  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleValProGlyLeu 120  
 DB 446 GCCAGTGAAGTCTCTCCAGTGAAGCTCTCCAGCAATGCAATCCATTAACCAAGAGCTT 505  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuValLeu 140  
 DB 506 GCTTTAAGCTGGAGAGTGGCTTATCATTTAGCTTACAGCTGTGAGACAGTTAAAAAG 565

QY 141 ProPheileProileHSHiMeGluAhiAalaleuThrIleargLeuThrasnLys 160  
 Db 566 CCATTTCATCCCATTCATCATATGAGGCTCATGCACTTACTATTAGTTCACCAATAA 625  
 QY 161 ValGluPheProPheLeuValleuLeuIleSerGlyGlyHisCysLeuLeuAaleuVal 180  
 Db 626 GTAGAAATTTCTTTTGTGTTCTTTGATTCTGGAGGTCACGTCTGTGTGACATTAGTT 685  
 QY 181 GluGlyValSerAppPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
 Db 686 CAAGAGATTTCAGATTTCCTGCTTCTGAAAGCTTTGGACATGACACAGGTGACATG 745  
 QY 201 LeuAspLysValAlaargArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220  
 Db 746 CTTGACAAAGGTGGAGAAAGACTTTCTTAAATAAACAATCCAGATGCTCCACCATGACT 805  
 QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGluGlyAsnArgPheHisPheAspIleLys 240  
 Db 806 GGTGGGAAAGCCATAGAATCATTTGGCCAAACAGAAATAGATTTCATTTGACATCAAA 865  
 QY 241 ProProLeuHSHiSAlaLysAsnCysAspPheSerPheThrGlyLeuGluHisValThr 260  
 Db 866 CCTCCCTTGACATGCTAAATAATGTGATTTTCTTTTACTGACCTTCACACACGTTACT 925  
 QY 261 AspLysIleIleMetLysGlyLysGluLysGluGlyIleGluLysGlyGluIleLeuSer 280  
 Db 926 GATTAATAATATGAAAACAGAAACAGAAAGAGATTTGAGAAAGGGCAATCTGTCT 985  
 QY 281 SerAlaAlaAspIleAlaIleAlaThrValGluHisPheMetAlaCysHisLeuValLysArg 300  
 Db 986 TCAGCAGACAGATTTGCTGCCACAGTACAGACACAAATGGAGATGTCATCTGTGAAGA 1045  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGluAsnArgPheLeuProGluAsnAsnAlaVal 320  
 Db 1046 ACAATCGGGCTATCTGTTTGTGTAAGCAGAGACTTGTAACTCAAAATATATCAGTA 1105  
 QY 321 LeuValAlaSerGlyValAlaSerAsnPheTrpIleArgArgAlaLeuGluIleLeu 340  
 Db 1106 CTGCTGATCTGGTGGTGGTGCAGAGTACTTATATCCGAGAGCTCTGGAAATTTTA 1165  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
 Db 1166 ACAAAACCAACACAGTGCATCTTGTGTCTCTCTCCACAGCTATGACATATATGCG 1225  
 QY 361 IleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 Db 1226 ATTATGATTCATGGAATGTGATTGAAGACTACGTGCTGGCATTTTACATGAC 1285  
 QY 381 IleGluGlyIleArgTrpGluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
 Db 1286 ATGGAAGGCAATCCCTATGAAACAAATAGTCTCTTGGAGTAGACATATCAAAAGAGTT 1345  
 QY 401 GlyGluAlaSerLysLysValProGluLeuLysMetGluIle 414  
 Db 1346 GGAGAGCTTCATTAATAGTCCACATTAATAATGAGATA 1387  
 RESULT 6  
 ABS57020  
 ID ABS57020 standard; cDNA; 1821 BP.  
 AC ABS57020;  
 XX  
 XX 30-JAN-2003 (first entry)  
 XX  
 XX cDNA encoding novel human glycoprotease 28472.  
 XX  
 XX Cancer; aberrant cell proliferation; aberrant cell differentiation;  
 XX breast cancer; ovarian cancer; prostate cancer; colon cancer;  
 XX lung cancer; immune disorder; heart disorder; cardiovascular disorder;  
 XX endothelial disorder; hematopoietic disorder; blood vessel disorder;  
 XX brain disorder; pain; metabolic disorder; liver disorder; diabetes;  
 XX platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;

KW autoimmune disorder; hypertension; atherosclerosis; heart failure;  
 KW myocardial infarction; ischaemic heart disease; Crohn's disease;  
 KW Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;  
 KW cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;  
 KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 147..1391  
 FT /tag= a  
 FT /product= "Glycoprotease 28472"  
 FT /note= "Specifically claimed in claim 1"  
 PN MO200277233-A2.  
 PD 03-OCT-2002.  
 PF 08-NOV-2001; 2001WO-US046724.  
 PP 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 PI Leiby KR, Kapeller-Libermann R, Glucksmann M;  
 XX WPI; 2003-029938/02.  
 DR P-PSDB; ABG71162.  
 DQ  
 XX  
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain  
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,  
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or  
 PT hypertension.  
 PS  
 XX  
 PS Claim 2: Fig 8A-B; 178bp; English.  
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and  
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The  
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine  
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding  
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise  
 CC sequences that encode a human seven transmembrane domain (7TM). The  
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide  
 CC sequences are useful for diagnosing, preventing or treating a subject  
 CC with or at risk of developing a disorder, e.g. cancer or aberrant  
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,  
 CC prostate, colon or lung cancer), immune disorders, heart disorders,  
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,  
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,  
 CC liver disorders or platelet disorders. These disorders include carcinoma,  
 CC sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,  
 CC hypertension, atherosclerosis, heart failure, myocardial infarction,  
 CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki  
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral  
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,  
 CC cachexia or diabetes. This sequence encodes the novel human glycoprotease  
 CC 28472  
 XX  
 SQ Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;  
 Alignment Scores:  
 Pred. No.: 1,24e-208 Length: 1821  
 Score: 2088.00 Matches: 407  
 Percent Similarity: 99.03% Conservative: 3  
 Best Local Similarity: 98.31% Mismatches: 4  
 Query Match: 98.26% Indels: 0  
 DB: Gaps: 0  
 US-10-649-273-2 (1-414) x ABS57020 (1-1821)  
 QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20

```

Db 147 ATGCTATCTGACCTAAGACGACGAGAGTTTAAAAACATCAAAAGAAAGTTAT 206
Qy 21 GIupheleuAgsSerPheAsnPhenHisProGlyThrLeuPheLeuHisLeuValLeu 40
Db 207 GAATTTTAAAGATTTTAAATTTTTCATCTGGACACATATTTCTTCAATAAATATATTTG 266
Qy 41 GYIIEGIuThSerCyAspAspThrAlaAlaValAlaAspGlyThrGlyValAsnVal 60
Db 267 GGAATTTGAAATTAAGTTGATGATGATACAGCAGCTGCTGGTGGAATGAAATGTG 326
Qy 61 LeuGIyGluAlaIleHisSerGlnThrGlyValHisLeuYsThrGlyGlyIleValPro 80
Db 327 TTGGAGAGACAAATACATTTCCAAACTGAAGTTGATTTAAAAACGGTGGATTTGCTT 386
Qy 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 387 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATAGTACAAAGAGCTTTCT 446
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
Db 447 GCCAGTGAAGTCTCTCCAGTGAACCTTCAGCAATTCGCACTACATTAACCAAGACTT 506
Qy 121 AlaLeuSerLeuGIyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuYsLys 140
Db 507 GCTTTAAGCCTGGAGTGGGCTTATCATTTAGCTTACAGCTGGTAGACAGTTAAAAAG 566
Qy 141 ProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 567 CCATTTCATTCCTCATTCATATAGAGGCTCATGACTTACTATTAAGTTAGTACCAATATA 626
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCyAspLeuLeuAlaLeuVal 180
Db 627 GTAGAAATTTCTTTTATGTTTATGTTTATGAGGCTGCTGCTGTTGCACTTAGT 686
Qy 181 GlnGlyValSerAspPheLeuLeuLeuGIyLysSerLeuAspIleAlaProGlyAspMet 200
Db 687 CAAGAGTTTCAGATTTCTGCTTCTTGAAGATCTTTGGACATAGCACAGTGAACATG 746
Qy 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCySerThrMetSer 220
Db 747 CTTCACAAAGTGGCAGAAAGACTTTCTTTAATAAATCCAGAGTGTCCACATAGAGT 806
Qy 221 GLYGLYysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 807 GGTGGAAAGCCATAGAAACATTTGGCCAAACAGAAATAGATTTCAATTTAACAATCAA 866
Qy 241 ProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 867 CTTCCCTTGCACTCATGCTTAAATATGATTTTCTTTTACTGGACTTCACACAGTTACT 926
Qy 261 AspLysIleIleMetLysLysGlyLysGlnGlyIleGlnLysGlyGlnIleLeuSer 280
Db 927 GATAAATAATATGAAACACAGAAACAGAGGAAAGATTTGAGAAAGGGCAATTCCTGCT 986
Qy 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCyHisLeuValLysArg 300
Db 987 TCAGCAGACGACATTTGCTGCCACAGTACAGCACAAATGGCAGTCACTTTGTAAAGAA 1046
Qy 301 ThrHisArgAlaIleLeuPheCyAspLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
Db 1047 ACACATCGGGCTATTTCTGTTTGTAAAGACAGAGACTGTTACCTCAAAATATATGACAGTA 1106
Qy 321 LeuValAlaSerGlyGlyValAlaSerAsnPhenYrIleArgArgAlaLeuGluIleLeu 340
Db 1107 CTGGTTGCATCTGGTGGTGGTGGCAAGTAACTTATATCCGACAGCTCTGGAAATTTTA 1166
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCyProProProArgLeuLeuCysThrAspAsnGly 360
Db 1167 ACAACAGCAACACAGTGCATTTGTTGTTCTCTCCACAGACTATGCACTATATATGAC 1226
Qy 361 IleMetIleAlaTPAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380

```

```

Db 1227 ATTATGATTCAGTGAATGATTTGAAGACTACGTCGCTGGCTTTTACATGAC 1286
Qy 381 IIEGIuGlyIleArgThrGlyProLysCyProLeuGIyValAspIleSerLysGluVal 400
Db 1287 ATGAAAGGATTCGGCTATGAAACCAAAATGTCCTTGGAGTACATATCAAAAGAAATT 1346
Qy 401 GLYGLUAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db 1347 GGAAGAGCTTCCATTAAGTACACAAATTAATAATGAGATA 1388

RESULT 7
ADA52832
ID ADA52832 standard; cDNA; 2208 BP.
XX
XX ADA52832;
AC
XX
XX 20-NOV-2003 (first entry)
DT
XX
XX Human coding sequence, SEQ ID 400.
DE
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
OS
XX Homo sapiens.
XX
XX EP1293569-A2.
PN
XX
XX 19-MAR-2003.
PD
XX
XX 21-MAR-2002; 2002EP-00006586.
PF
XX
XX 14-SEP-2001; 2001JP-00328381.
PR
XX 24-JAN-2002; 2002US-0350435P.
PR
XX
XX (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX
XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX
XX WPI: 2003-395539/38.
DR P-PSDB; ADA54471.
XX
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
XX Claim 1; SEQ ID NO 400; 205pp; English.
PS
XX
XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
XX Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;
SQ

Alignment Scores:
Pred. No.: 2,15e-193 Length: 2208
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: 10 Gaps: 1

US-10-649-273-2 (1-414) x ADA52832 (1-2208)
Qy 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
Db 344 ATCTATATCTTGACTAAGACTGCAGAGGTTTTTTTTTAAACATCAAAAGAAAGTTTAT 403

```

```

QY 21 GIuPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLeuValLeu 40
Db 404 GAAATTTTAAAGAAATTTTAAATTTTCATCTGGACACATATTTCTTCAATAATATGATGG 463
QY 41 G1Y1LeGIuThrSerCyAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60
Db 464 GGAATTTAAACTAGTTGATGATATACAGACAGCTGCTGTGTGTGAATGAACCTGAAATGCG 523
QY 61 LeuGIyGIuAlaIleHisSerGlnThrGluValHisLeuLeuThrGlyGlyIleValPro 80
Db 524 TTGGGAGAAGAAATATACATTTCCCAAACTGAATTCATTTTAAAAACAGTGGGATTTGCTCT 583
QY 81 ProAlaIaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100
Db 584 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCACCAATATGATCAAGAACTCTTTCT 643
QY 101 AlaSerGIyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120
Db 644 GCCAGTGAAGCTCTCCCAAGTACCTTCACCAATTTGCAACATTAACCAATMAACAGACTT 703
QY 121 AlaLeuSerLeuGIyValGIyLeuSerPheSerLeuGlnLeuValGIyGlnLeuLeuLys 140
Db 704 GCTTTAAGCTCGGAGTGGCTTTATCATTTAGCTTACAGCTGTGTAGACATTTAAATAAG 763
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 764 CCAATTCATTCATTCATCATATGAGGCTCATGACATTAAGTTAGGTTAGCCATATAA 823
QY 161 ValGIuPheProPheLeuValLeuLeuIleSerGIyGlyHisCyLeuLeuAlaLeuVal 180
Db 824 GTAGAAATTTCTTTTATGATCTTTTGTGATTTCTGAGGCTCACTGTGTGGCATTAAGTT 883
QY 181 GlnGIyValSerAspPheLeuLeuLeuGIyLysSerLeuAspIleAlaProGlyLysMet 200
Db 884 CAAAGAGTTTCAGATTTTCTGCTTCTTGAAAGCTTTTGACATAGCACCAAGTATACATG 943
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyLysSerThrMetSer 220
Db 944 CTGGACAAAGGTGGCAAGACCTTCCTTATTAATAACATCCAGAGTCTCCACCATGAGT 1003
QY 221 GlyGIyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 1004 GGTGGGAAACCCATAGAAACATTTGGCCAAACAGAAATATGATTTCAATTTGACATCAAA 1063
QY 241 ProProLeuHisHisAlaLysAsnCyAspAspPheThrGlyLeuGlnHisValThr 260
Db 1064 CCTCCCTTGACATGCTGCTAAATAATGTGATTTTCTTTTACGTGACCTTCAACACCTTACT 1123
QY 261 AspLysIleIleMetLysLysGluLysGluGlyIleGlnLysGlyGlnIleLeuSer 280
Db 1124 GATTAATTAATTAATGAAGAAAGAAAGAAAGGATTTAGAAAGGGCCAAATCCTGTCT 1183
QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCyHisIleLeuValLysArg 300
Db 1184 TCAGCAGCAGACATGCTGCGACAGTACAGACACAAATGGCATGTCCTGTGTAAAGA 1243
QY 301 ThrHisArgAlaIleLeuPheCyLeuGlnArgAspLeuLeuProGlnAsnAlaVal 320
Db 1244 ACACATCGGGCTATTCGTGTTTGTAGCAGAGACATTTGACCTCAAAATAATGACAGTA 1303
QY 321 LeuValAlaSerGIyValAlaIleSerAsnPheThrIleArgArgAlaLeuGlnIleLeu 340
Db 1304 CTGTGTGCATCTGTGTGTGTGCAAGTAATCTTGTAATCCGACAGCTCTGAAATTTTA 1363
QY 341 ThrAsnAlaThrGlnCyThrLeuLeuCyAspProProAlaGlyLeuCyThrAspAsnGly 360
Db 1364 ACAACCCACACAGTGCACATTTGTGTGTCTCTCCCAACATATGACATGATATGAC 1423
QY 361 IleMetIleAlaThrPasnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1424 ATTATGATGCA----- 1435
QY 381 IleGluGIyLysArgGlyGluProLysCyAspProLeuGIyValAspIleSerLysGluVal 400

```

```

Db 1436 -----TCATGTCTCTTGGAGTACACATATCAAAAGAGTT 1471
QY 401 GIyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db 1472 GGAAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1513

RESULT 8
ADQ24627
ID ADQ24627 standard; DNA; 2890 BP.
XX
AC ADQ24627;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnick A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 7447; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;

Alignment Scores:
Prid. No.: 3,2e-193 Length: 2890
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Best Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: 12 Gaps: 1

US-10-649-273-2 (1-414) x ADQ24627 (1-2890)
QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
Db 1001 ATGCTAATCTTGACTAAGACTGACAGAGTTTATTTTAAACATCAAAAGAAAGATTAT 1060

```

QY 21 GluPheLeuArgSerPheAspPheHisProGlyThrLeuPheLeuHisIleValLeu 40  
 DB 1061 GAATTTTAAAGAACTTTTAAATTTTCATCTCGAACAACATATTTCTCATTAATAATGATATTTG 1120  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValIleAspGluThrGlyAsnVal 60  
 DB 1121 GGAAATTAAGATAGTGTGAATGATACAGACACTGCTGTGGTGAATGAACCTGGAAATATGTG 1180  
 QY 61 LeuGlyGluAlaIleHisSerGluThrGlyValHisLeuLysThrGlyIleValPro 80  
 DB 1181 TTGGGAGAGCAATACATTTCCCAACTGACAGTTCATTTAAAACAGGTGGGATTTTCTCT 1240  
 QY 81 ProAlaIleGluGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 1241 CCACGAGCTCAACAGCTTCACAGAGAAATATTCACAAATATGATCAAGAAAGCTCTTTCT 1300  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrIleLeuProGlyLeu 120  
 DB 1301 GCCAGTGAAGTCTCTCCAAAGTACCTTCACCAATTTGCACTACCATTAACACAGACTT 1360  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleLeuLys 140  
 DB 1361 GCTTTAAAGCTGGAGTGGGCTTATTCATTAGCTTACAGCTGTGAACAGTTAAATAAG 1420  
 QY 141 ProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
 DB 1421 CCATTCAATTCATTCAATCATATGAGGCTCATGCACTTACTATAGGTTGACCAATAAA 1480  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 DB 1481 GTAAATAATTCCTTTTATGATCTTTTGAATTTCTGAGAGTCACTGTCTGTGGCATTAAGT 1540  
 QY 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyLysPhe 200  
 DB 1541 CAAGGAGTTTCAAGATTTTCTGCTTGTGAAAGCTTTTGACATATGACACAGTATCAATG 1600  
 QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyLysSerThrMetSer 220  
 DB 1601 CTTGACAAAGGTGGCAAGAAAGCTTCCTTATTAATAACATCCAGAGTCTCCACCATAGGT 1660  
 QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 DB 1661 GGTGGGAAGCCATAGAAACATTTGGCCAAACAGAAATTAAGATTCATTTGACATCAAA 1720  
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisAlaThr 260  
 DB 1721 CTTCCCTTGCAATGCTTAAATAATTTGATTTTCTTTTACTGAGCTTCAACACCTTACT 1780  
 QY 261 AspLysIleIleMetLysLysGluLysGluGlyIleGluLysGlyIleLeuSer 280  
 DB 1781 GATTAATAATTAATGAAAGAAAGAAAGAAAGGATTTGAGAGGGGCAAAATCTGTCT 1840  
 QY 281 SerAlaIleAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
 DB 1841 TCACGACAGACATTTGCTGCCACAGTACAGACACAAATGAGATGTCATCTGTGAAAAGA 1900  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaVal 320  
 DB 1901 ACACATGGGGCTATTCGTTTGTGTAAGCAGAGACTTGTAACTCCAAATAATATCCAGTA 1960  
 QY 321 LeuValAlaSerGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeu 340  
 DB 1961 CTGCTTGATGATGGTGGTGGTGCAGAGTAATCTGTATCCGAGAGCTCTGCAAAATTTTA 2020  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
 DB 2021 ACAAAACCAACAGAGTGAACCTTTGTGTGCTCTCTCCACAGATATGACATATATGAC 2080  
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 2081 ATTTATGATTTGCA----- 2092  
 QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerIleGluVal 400

DB 2093 -----TCATGTCCTTGTGAGTGAACATATCAAAAGAGTT 2128  
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 DB 2129 GGAGAGCTTCCATTAAGTACCAATTAATAATGAGAGATA 2170  
 RESULT 9  
 ABX70950  
 ID ABX70950 standard; cDNA; 1416 BP.  
 XX  
 AC ABX70950;  
 XX  
 DT 05-MAR-2003 (first entry)  
 XX  
 DE Novel human cDNA sequence #175.  
 KW Human; gene; ss; nervous system disorder; peripheral neuropathy;  
 KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;  
 KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;  
 KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;  
 KW insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;  
 KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;  
 KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;  
 KW coagulation disorder; cancer; tumour; inflammatory disease; septic shock;  
 KW Crohn's disease; anaphylaxis; proliferation; chemotactic;  
 KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;  
 KW haemostatic; antiinflammatory; expressed sequence tag; EST.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W020281731-A2.  
 PD 17-OCT-2002.  
 XX  
 PF 29-JAN-2002; 2002W0-US001222.  
 XX  
 PR 30-JAN-2001; 2001US-00774528.  
 XX  
 PA (HSE-) HSEQ INC.  
 XX (GOOD/) GOODRICH R W.  
 PI Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QN, Ren F;  
 XX Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;  
 DR WPI; 2003-058563/05.  
 XX  
 PT Novel polypeptide useful for treating neurodegenerative diseases, myeloid  
 PT or lymphoid cell disorders, bone disorders, mechanical and traumatic  
 PT disorders, coagulation disorders, and inflammatory diseases.  
 XX  
 PS Claim 1; Page; 612pp; English.  
 XX  
 CC This invention relates to the cDNA sequences encoding an isolated novel  
 CC human polypeptide. The protein encoded by the nucleic acid of the  
 CC invention is useful for treating central and peripheral nervous system  
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic  
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,  
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus  
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)  
 CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia)  
 CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,  
 CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head  
 CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;  
 CC bacterial, viral or fungal infections; allergic conditions such as  
 CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);  
 CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's  
 CC disease, anaphylaxis). The protein may be used to inhibit the growth,  
 CC infection or function of infectious agents such as bacteria, fungi,  
 CC viruses, or to effect bodily characteristics, biorhythms or circadian  
 CC cycles of rhythms. The protein may also have  
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis  
 CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,

CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
CC activities. The cDNA sequences of the invention are useful for expressing  
CC recombinant protein for analysis. The present sequence represents a novel  
CC human cDNA sequence of the invention, this sequence is an expressed  
CC sequence tag (EST) and was identified using subtractive hybridisation

XX Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	5,31e-173	Length:	1416
Score:	1747.00	Matches:	340
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.71%	Mismatches:	0
Query Match:	82.21%	Indels:	0
	8	Gaps:	0

US-10-649-273-2 (1-414) x ABX70950 (1-1416)

```

QY 74 LysThrGlyGlyIleValProProAlaAgiGlnGlnLeuHisArgGluAsnIleGlnArg 93
Db 280 AGAACAGGTGGATTTCTTCACAGCAGCTCAACAGCTTCAAGAAATTTCAAGA 339
QY 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113
Db 340 ATACTACAGAAGCTCTTCTGCCAGTGCAGTCTCCAGAGTCACTCTCAACAATTGCA 399
QY 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133
Db 400 ACTACCAATAAACCGAGCTTGAAGCCGGAGGAGGCTTATCATTTAGCTTACG 459
QY 134 LeuValGlyGlnLeuLysLeuProPheIleProIleHisHisMetGluAlaHisAlaLeu 153
Db 460 CTGATGAGCAGCTTAAABAAAGCCATTCCCATTCATCATGAGGCTCATGACTT 519
QY 154 ThrIleArgLeuThrAsnLysValGlnPheProPheLeuValLeuLeuIleSerGly 173
Db 520 ACTATTAGTTGACCAATAAAGTAGAATTTCTTTTATTTGTTTCTTGAGGT 579
QY 174 HisGlyLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193
Db 580 CACTGTCTGTTGGATTAGTTCAGAGAGTTCAGATTTCTGCTCTTGGAAGCTTTG 639
QY 194 AspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleHis 213
Db 640 GACATGACCAAGTGCATGCTTGCACAGGTGGCAAGAGCTTTCTTAATAAAGCT 699
QY 214 ProGluCysSerThrMetSerGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsn 233
Db 700 CCAGAGTCTCCACCATGAGTGGTGGAAAGCCATAGAAACATTGGCCAAACAGAAAT 759
QY 234 ArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPhe 253
Db 760 AGATTTCATTTGACATCAACCTCCCTGCATCATGCTAAATAATGATTTTCTTT 819
QY 254 ThrGlyLeuGlnHisValThrAspLysIleIleMetLysGluLysGlnGlnGlyLe 273
Db 820 ACTGAGCTTCAACAGTACTGATTAATAATTAAGAAAGGAAAGAGAGAGTAT 879
QY 274 GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMet 293
Db 880 GAGAAAGGCGAAATCTCTCTTCAGCAGCAGACATTCCTCCACAGTACAGCAGCAATG 939
QY 294 AlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnHisArgPhe 313
Db 940 GCATGTCATCTTGAAAGAAACATCGGCTATTTGTTTGAACAGAGACACTTG 999
QY 314 LeuProGlnHisAsnAlaValLeuValAlaSerGlyValAlaLysAsnPhePheIle 333
Db 1000 TTACTCTAATAATATATGAGTCTGTTGCATCTGTGTGTGTGCGCAAGTACTTATATTC 1059
QY 334 ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro 353
Db 1060 CGCAGAGCTCTGGAATTTTAACAAACGCAACAGTCACTTGTGTGTCTCTCC 1119

```

```

QY 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAla 373
Db 1120 AGACTATGCATCATATATGCAATTAATGATTCATGAGAAATGATTAAGAAACATCGTCT 1179
QY 374 GlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTrpGluProLysCysProLeuGly 393
Db 1180 GGCTTGGCATTTTATCATGACATGAGAGCATCGGCTATGACCAAAATGTCTCTTGA 1239
QY 394 ValAspIleSerIleGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGlu 413
Db 1240 GTAGACATATCAAAAGAAAGTTGAGAAAGCTTCATTAAGTACACAAATTAATGAG 1299
QY 414 Ile 414
Db 1300 ATG 1302

RESULT 10
AB075508
AB075508 standard; DNA; 1572 BP.
XX
XX
XX AB075508;
XX
XX 07-NOV-2002 (first entry)
XX
DE Murine sialoglycoprotease-like gene sequence SEQ ID NO.7.
XX
XX Murine, mouse; protease; calcium activated neutral protease type 5;
XX CAPNs; trypsin 4; sialoglycoprotease; enzyme; genetic disease;
XX neurological; neuropsychological; psychotic illness; transgenic animal;
XX gene; db.
XX
XX Mus musculus.
XX
XX WO200245491-A2.
XX
XX 13-JUN-2002.
XX
XX
XX 05-DEC-2001; 2001WO-US046405.
XX
XX 06-DEC-2000; 2000US-0251803P.
XX PR 06-DEC-2000; 2000US-0251820P.
XX PR 13-DEC-2000; 2000US-0255971P.
XX
XX (DELT-) DELTAGEN INC.
XX
XX Allen KD, Leviten MW;
XX
XX WPI; 2002-657389/70.
XX
XX Novel transgenic animal, comprising a disruption in protease target gene,
XX is useful for identifying agents that ameliorates a phenotype associated
XX with a disruption in a protease target gene.
XX
XX Example 3; Fig 7; 62pp; English.
XX
XX The present invention describes a non-human transgenic animal (I)
XX comprising a disruption in a protease target gene (PG) selected from
XX calcium activated neutral protease type 5 (CAPNs) gene, trypsin 4 gene
XX and sialoglycoprotease-like gene. Also described is a targeting construct
XX (II), comprising a first polynucleotide sequence homologous to at least a
XX first portion of PG, a second polynucleotide sequence homologous to at
XX least a second portion of PG and a selectable marker. (II) is useful for
XX producing a transgenic mouse comprising a disruption in a protease target
XX gene, by introducing (II) into a cell, introducing the cell into a
XX blastocyst, implanting the resulting blastocyst into a pseudopregnant
XX mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
XX and breeding the chimeric mouse to produce the transgenic mouse. (I) is
XX useful for identifying an agent that modulates the expression or function
XX of a protease target gene, by administering an agent to (I) and
XX determining whether the expression or function of the disrupted protease
XX target gene in (I) is modulated. (I) is also useful for testing the
XX efficacy of proposed genetic and pharmacological therapies for human

```

CC genetic diseases, such as neurological, neuropsychological or psychotic  
 CC illness. The present sequence represents murine sialolipoprotease-like  
 CC gene sequence, which is used in an example from the present invention

XX Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;

# Alignment Scores:

Pred. No.:	3.76e-152	Length:	1572
Score:	1549.00	Matches:	326
Percent Similarity:	84.50%	Conservative:	23
Best Local Similarity:	78.93%	Mismatches:	60
Query Match:	72.89%	Indels:	7
		Gaps:	2

US-10-649-273-2 (1-414) x ABQ75508 (1-1572)

```

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
DB 5 ATGCTAATGTTTAAAGAGAGACAGACGACCTATTCCCAAGCCCCCAAGAGTAAAGTTTAT 64
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysLysIleValLeu 40
DB 65 GGAATTTTAAAGAAAGATTGATGTTGATCCAGAACTCTCTCTTGTCAATAACTGGATCTG 124
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaAlaValAlaAspGluThrGlyAsnVal 60
DB 125 GGAATTTGAAACCACTGATGATGACACAGAGCGCTGTGTGATGAACTGGAAATGTG 184
QY 61 LeuGlyAlaAlaIleHisSerGlnThrGluValHisLysLysThrGlyIleValPro 80
DB 185 CTGGGGAGAGACATGCACTCCCAAACTCAGGTTTCATCGAATAAGGGGAGATTCTTCT 244
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnAlaGlyIleValGlnAlaLeuSer 100
DB 245 CCAGTACTCTCAACAATTCTCAGAGAAAATATTCACACAAATAGTAAAGAAATCTTTCT 304
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120
DB 305 GCCGTAGAAATCACCCCAAGGATCTCTCAGCAATTCGCACTACCAACCGGAGACTG 364
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleLeuLysLys 140
DB 365 GCCCTAAGTTTGGAGTTGGCTTATCCTTTGCTTACAGCTAGTAAATCAGTTTAAARS 424
QY 141 ProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
DB 425 CCATTCAT-CCGATTCATCACAATGAGGCTCAGCAGACATATTAAGGCTCCACAAATAA 483
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
DB 484 GTAGATTN-CTTTTATGTTCTTGTGATTTCTGGGGCTGCTGCTGTGAMRTTAGTC 542
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyLysPhe 200
DB 543 CAAAGTGTTCCTGATTCCTGCTCTGGAAGCTTTTGGACATAGCAGCGGCACTATG 602
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlnCysSerThrMetSer 220
DB 603 CTTCACAAAGGTGGCAGAGAAAGCTTTCTTAAATCAAAACATCCAGAAATGTTTCAATGAGT 662
QY 221 GlyIleLysValAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB 663 GGTGGAAGGCTAATAGAACAGTGTGGCAAGACGAAATAGATTCATTATTATCTATCAT 722
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
DB 723 CCACCTATGCAAGATGCTAAGAAATTCGCAATTTTCTTTCACGGGACTTCACATATTAAT 782
QY 261 AspLysIleIleMetLysLysGlyLysGlyLysGlyIleGlyLysGlyIleLeuSer 280
DB 783 GATTAAGCTAATTAACACCAAGAAAGAAAGAAAGAGGCAATTAAGGCAAAATCTGTCA 842
QY 281 SerAlaAlaSerIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300

```

```

DB 843 TAGGTGAGACATTGCTCTGCTCGGTACAGATGCAACAGCGTGCACCTTGCGAAAG 902
QY 301 ThrHisArgAlaIleLeuPheCysLysGlnAlaGAspLeuLeuProGlnHisAsnAlaVal 320
DB 903 ACACATCGGCTAATCTGTGTTTGCAAGAGAAATTTGCTCTTCACACTAAGCAGTA 962
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArgAlaLeuGluIleLeu 340
DB 963 TTAGTTGATCTGAGAGGTGTTCAGATTAATCTTGATCTCGAAGACATTTGAAATGTTC 1022
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
DB 1023 GCAAAATGCAACGACATGACAGGTTGTGTGTCCACT-TCAAGACTGTGCACTGCAATGGG 1081
QY 361 IleMetIleAlaITPAsnGlyIleGluValGluAlaGlyIleLeuHisAsn 380
DB 1082 CATATGATTTGCAATGGAATGGAATTCGACGATTAACGATTAACGATTAACGATTA 1141
QY 380 P-IleGluGlyIleArgGlyLysProLysCysProLeuGlyVal--AspIleSerLysG 399
DB 1142 TGAATGAGACATTCGGTTATTAACCAAAATGCTCTTGTGATGAGGCAATTATCCGAA 1201
QY 399 IuValGlyGluAlaSerIleLysValProGlnLeu 410
DB 1202 AGTTTGCGAGA---AGCTTGCCCATTAATAAGTTA 1233
RESULT 11
ABST76639 standard; DNA; 1526 BP.
AC ABST76639;
XX 11-DEC-2002 (first entry)
DE DNA encoding novel human metalloprotease MP1 fragment #1.
XX
KW Metalloprotease; MP-1; immune disorder; glucamate transport; cancer;
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;
KW neurological disorder; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200272751-A2.
XX
PD 19-SEP-2002.
XX
PF 05-FEB-2002; 2002WMO-US003353.
XX
PR 05-FEB-2001; 2001US-0266518P.
XX
PR 10-APR-2001; 2001US-0282814P.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Chen J, Feder J, Nelson TC, Duclos F, Kyvstek S,
XX
DR MPI, 2002-723329/78.
XX
DR P-PSDB; ABG96487.
XX
PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,
PT treating, or ameliorating diseases associated with aberrant
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and
PT neurological disorders.
XX
PS Disclosure; Page 462-463; 473pp; English.

```



CC The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
 CC ameliorating a medical condition, particularly an immune disorder, an  
 CC aberrant glutamate transport or motor neuron disorder, such as  
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
 CC condition. The compositions and methods are also useful for diagnosing,  
 CC prognosticating, treating, ameliorating and/or treating disorders  
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease). This sequence represents a  
 CC metalloprotease MPI polynucleotide

XX Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	5.9e-135	Length:	1526
Score:	1385.00	Matches:	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	65.18%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2 (1-414) x ABS76639 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGluPheProPheLeuVal 167  
 DB 1 ATGAGGCTCATGACATCTACTATTAGGTGACCAATAAGATGATTTCTTTTACTGTT 60  
 QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187  
 DB 61 CTTTTCATTTCTGAGAGTCACTGTCGTGGCATTAAGTTCAGAGTTTCAGATTTTCTG 120  
 QY 188 LeuLeuGlyIleValSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 207  
 DB 121 CTTCTTGGAAAGCTTTGGACATGACACAGAGTGCATGCTTGACAGGTGGCAAGAAAGA 180  
 QY 208 LeuSerLeuIleValSerProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 227  
 DB 181 CTTTCTTTAATAAATCCAGAGTCTCCACCATGAGTGGTGGAAAGCCATAGAACAT 240  
 QY 228 LeuAlaIleGlnGlyAsnArgPheHisPheAspIleValProProLeuHisHisAlaIle 247  
 DB 241 TTGGCCAAACAGAGAAATAGATTTTCATTTTGACATCAAACTCCCTTCATCATCTATAA 300  
 QY 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleValIleMetIleVal 267  
 DB 301 AATTTGATATTTCTTTACTGAGACTTCACACCGTTACTGATTAATAATATGAAAG 360  
 QY 268 GluIleGlnGlyIleGluIleValGlnIleLeuSerSerAlaAlaAspIleAlaIle 287  
 DB 361 GAAAAAAGAGAGGATTTGAGAAAGGGGCAAAATCCGTCTTACGACGACATTCCTCC 420  
 QY 288 ThrValGlnHisThrMetAlaCysHisLeuValIleValArgThrHisArgAlaIleLeuPhe 307  
 DB 421 ACAGTACAGACACAAATGCGATTCATCTTGTGAAAAGAACATCGGGCTATTTCTGTTT 480  
 QY 308 CysIleGlnArgAspLeuLeuProGlnAsnAsnAlaValIleValAlaSerGlyVal 327  
 DB 481 TGTAAAGAGAGAGACTGTTAATCTCAAAATATGACAGTACGTGTCGTCGTC 540  
 QY 328 AlaSerAspPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 347  
 DB 541 GCAAGTAACTTCTATATCCGACAGCTCTGGAATTTTMAAAACGCAACACAGTGCAT 600  
 QY 348 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleTrpAsnGly 367  
 DB 601 TTGTTGTCTCTCTCCCAAGCTATGCACTGATTAATGCAATTAAGATTCAGATGAGATGT 660

QY 368 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgGly 387  
 DB 661 ATTGAAGAGCTACGCTCTGCTGGCATTTTACATGACATAGACATCCGCTATGAA 720  
 QY 388 ProValCysProLeuGlyValAspIleSerValGluValGlyAlaSerIleVal 407  
 DB 721 CCAAAATGCTCTTGGAGTACATATCAAAAGAGTTGAGAGAGCTTCATTAAGTA 780  
 QY 408 ProGlnLeuIleMetGlnIle 414  
 DB 781 CCACATTTAAATGAGAGATA 801  
 RESULT 12  
 ADE31345/C  
 ID ADE31345 standard; DNA; 3358 BP.  
 XX  
 AC ADE31345;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human diagnostic and therapeutic polynucleotide (dthp), SEQ ID No 100.  
 XX  
 KW diagnostic and therapeutic polynucleotide; dthp; antiarteriosclerotic;  
 KW antiinflammatory; cerebroprotective; antilipemic; antidiabetic;  
 KW immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;  
 KW osteopathic; antiarthritic; antineumatic; cyostatic; hepatotropic;  
 KW virucide; haemostatic; anti-HIV; antithyroid; thyromimetic;  
 KW dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;  
 KW thrombolytic; anticoagulant; anorectic; vasotropic; antilucer;  
 KW gene therapy; protein replacement therapy; human; gene; ds.  
 XX  
 OS Homo sapiens.  
 PN WO2003062376-A2.  
 XX  
 PD 31-JUL-2003.  
 XX  
 PF 13-JAN-2003; 2003WO-US001096.  
 XX  
 PR 16-JAN-2002; 2002US-0349384P.  
 XX  
 PR 17-JAN-2002; 2002US-0349413P.  
 XX  
 PR 17-JAN-2002; 2002US-0349946P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GB, Jackson JL,  
 PI Yu JY, Tusson O, Yap PE, Ameshey SR, Dam TC, Liu TP, Gerstin EH;  
 PI Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RV, Uraehka ME;  
 PI Krieknam SR, Kolluru V, Panesar IS;  
 XX  
 DR WPI; 2003-636732/60.  
 DR P-PSDB; ADE31156.  
 XX  
 PT New human diagnostic and therapeutic polynucleotides and polypeptides,  
 PT useful for diagnosing, treating or preventing e.g. leukemia, brain  
 PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke  
 PT or Alzheimer's.  
 XX  
 PS Claim 1; SEQ ID NO 100; 634pp; English.  
 XX  
 CC The invention relates to a novel isolated human diagnostic and  
 CC therapeutic polynucleotide (designated dthp). The novel dthp  
 CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798  
 CC base pairs fully defined in the specification; a polynucleotide  
 CC comprising a naturally occurring polynucleotide sequence at least 90%  
 CC identical to the dthp polynucleotide; a polynucleotide complementary to  
 CC the dthp polynucleotide or its polynucleotide which is at least 90%  
 CC identical; or an RNA equivalent of any of the polynucleotides mentioned  
 CC above. The dthp polynucleotides have the following activities:  
 CC antiarteriosclerotic, antiinflammatory, cerebroprotective, antilipemic,  
 CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,  
 CC tranquilizer, osteopathic, antiarthritic, antineumatic, cyostatic,



CC hepatotropic, virucide, haemostatic, anti-HIV, antithyroid, thyromimetic,  
 CC dermatological, antibacterial, fungicide, antiparasitic, anticonvulsant,  
 CC thrombolytic, anticoagulant, anorectic, vasotonic, and anticancer. The  
 CC novel DHP polynucleotides polypeptide can be used in gene therapy and  
 CC protein replacement therapy. The dHP polynucleotides or DHP  
 CC polypeptides are useful for diagnosing, preventing or treating diseases  
 CC associated with the expression of human molecules. In particular, these  
 CC diseases include cancers (e.g. adenocarcinoma, leukemia, melanoma, brain  
 CC cancer), breast cancer, cervix cancer, bone cancer, liver cancer, lung  
 CC cancer) or other cell proliferative disorders (e.g. arteriosclerosis,  
 CC atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary  
 CC thrombocytopoenia), autoimmune/inflammatory disorders (e.g. AIDS,  
 CC Addison's disease, thyroiditis, Crohn's disease, Graves' disease,  
 CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid  
 CC arthritis, scleroderma, systemic lupus erythematosus), infectious (e.g.  
 CC viral, bacterial, fungal or parasitic infection), developmental disorders  
 CC (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.  
 CC chromoblastoma, hypopituitarism, hypogonadism, gigantism, goiter) metabolic  
 CC disorders (e.g. hypercholesterolemia, hypoglycaemia, diabetes,  
 CC hyperlipidaemia, obesity), neurological disorders (e.g. ischemic  
 CC cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,  
 CC Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,  
 CC anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),  
 CC transport disorders (e.g. akinesia or multidrug resistance), or  
 CC connective tissue disorders (e.g. Paget's disease or rickets). This  
 CC polynucleotide sequence represents one of the human dHP DNA sequences  
 CC of the invention.

XX Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.:	339e-94	Length:	3358
Score:	1001.50	Matches:	250
Percent Similarity:	44.82%	Conservative:	1
Best Local Similarity:	44.64%	Mismatches:	7
Query Match:	47.13%	Indels:	307
		Gaps:	2

US-10-649-273-2 (1-414) x ADE31345 (1-3358)

```

QY 157 LeuThrsValAlaArgPheProPheLeuValLeuLeuLeuSerGlyValHisCysLeu 176
DB 3067 GTGACCAATAAGAGAGATTCCTTTTAACTCT-TGATTTCTGGAGGTCACTGTCG 3009
QY 177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAsp11eala 196
DB 3008 TGGGCATTAGTTCAGAGAGTTCAGATTCTGCTTGGAAAGTC-TTGGACATAGCA 2950
QY 197 ProGlyAspMet----- 200
DB 2949 CCAAGTGCATGCGCTTGAACAAAGTAATTAAAGATTAATTTCTCCATCTTTTGTATGT 2890
QY 200 ----- 200
DB 2889 TGTCCATTTCACCTAAGTAGCAATAGATGTGTACACACATTCACCTAAATATTTCTGAA 2830
QY 200 ----- 200
DB 2829 TTTATCTTAGTAAGTGAACAAATAATTCACATATGTTGGAGAAAAATAGAAAGCAGTA 2770
QY 200 ----- 200
DB 2769 GTACACGAATTTAATTAATCTTACCTTCTTAATAAATGTAAGAGTTCAATCTGTA 2710
QY 200 ----- 200
DB 2709 CATAAAGCGTAATAGTTTGCAAGTACAGTTATGATTTTGCACAAATAATATATGTGAA 2650
QY 200 ----- 200
DB 2649 AGAAGTGTCTCGTAACCTAATACATACGCAAAAAAGTAAATAGAGAAATATATATAGA 2590
QY 200 ----- 200

```

```

DB 2589 TTAACATAAGACATTAAGATGCAATGACAGAAATTAATCAACAATTAATTACACCA 2530
QY 200 ----- 200
DB 2529 CAGACAGGTCCCGCCGACCCCTTTGTTTGAATTAATAAGAGGTACTGCAATAT 2470
QY 200 ----- 200
DB 2469 ATAGAAACTACAAACAACAACAACAGACAGTCTTCCACAGTGAATTAATAGAAATAGG 2410
QY 200 ----- 200
DB 2409 ACAAGTTCTTATTAATTAATGACGTTCAATCATTAAGACGTTATGTCACTTCAAGCCATTT 2350
QY 200 ----- 200
DB 2349 CCACCAATAGAAAGAGCAAAATAGACAGGAGCAGTATGACCTTATTTGCGGTC 2290
QY 200 ----- 200
DB 2289 ATCATTAAGACAGAGGTGTCTGCTTACTGAATATCAGCTATAGCTATATTTGCCAAA 2230
QY 200 ----- 200
DB 2229 GTATAGCATTTTATTAATTAATTCAGGGGTTTTTTGTTTGTAGTAATTTCAATTAATTT 2170
QY 201 -----Leu 201
DB 2169 CCTTGCATCTTTGCTTTCACAGTATTAATTAATTAATGACCTGAAAAATATGTTCTTT 2110
QY 202 AspLysValAlaArgArgLeuSerLeuLeuLysHisProGluCysSerThrMetSerGly 221
DB 2109 GATAG-GTGGCAAGAGACTTCTTTAATTAACATCCAGAGTCTTCCACATGAGTGGT 2051
QY 222 -GlyLysAlaLeuGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAsp11eLysPr 241
DB 2050 GGGAAAGCCATAGAACATTTTGGCCAAACAGAAATATGATTTCAATTTTGACATCAAC 1991
QY 241 oProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAs 261
DB 1990 TCCCTTGCACTCATGCTTAATAAATTTGATTTTCTTTTACTGACCTTCAACACGTTACTGA 1931
QY 261 PLYSILELLEMETLVBVSGULVSGULV----- 271
DB 1930 TAAATTAATTAATGAAGAAAGAAAGAGAGATATTTCTTAATTAATGAAGTGAACA 1871
QY 272 -----Gly11LeuLysG 276
DB 1870 GATTAATATTCCTGCAATGCTGCTTAATAATTAAGCTGCTCATTTCTGCAGGTATTGACAGG 1811
QY 276 IYGLNILEUSeSerAlaAlaAsp11eAlaAlaThrValGlnHisThrMetAlaCysH 296
DB 1810 GGCATATCTGATCTTCAGAGAGACATTTGTCGCCACATGACAGCACAAATGCAATGTC 1751
QY 296 ILeuValLysArgThrHisArgAlaAlaLeuPheCysLysGlnAsnArgAspLeuProG 316
DB 1750 ATCTTGTAAGAAACACATCGGGCTATTTCTTTTGTAGACAGAAAGCTGTGACCTC 1691
QY 316 IAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheYr11eArgArgA 336
DB 1690 AATAATATGCACTGCTGTTGCAATCTGT-GGTGTGCAAGTAATCTTATATCCGAGAG 1632
QY 336 ILeuGluL1LeuThrsAlaAlaThrGlnCysThrLeuLeuCysProProArgLeu 356
DB 1631 CTCTGGAATTTTAAAC-AAAGCAACAGTGAACCTTGTGTGCTCTCCCAAGCTAT 1573
QY 356 YThrAspAsnGly11eMet11eAlaTPAsnGly11eGluArgLeuAlaGlyLeu 376
DB 1572 GCACTGATATATGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1513
QY 376 IY11LeuHisAsp11eGluGly11eArgTyrGluProLysCysProLeuGlyValAsp1 396

```

Db 1512 GCATTTCATGACATGAGAGGCATCGCTATGAAACAAATGTCCTTGAGTAGACA 1453  
Qy 396 leSerLySGluValIGlyValaSerIleLyValProGlnLeuLyMetGluIle 414  
Db 1452 TATCAAAAGAAAGTTGAGACAGCTTCATTAAGTACCAATTAATAATGAGATA 1397  
RESULT 13  
ABL24633  
ID ABL24633 standard; DNA; 1601 BP.  
XX ABL24633;  
AC ABL24633;  
DT 26-MAR-2002 (first entry)  
DB Drosophila melanogaster genomic polynucleotide SEQ ID NO 25372.  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 25372.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical; gene; ds.  
XX Drosophila melanogaster.  
OS Drosophila melanogaster.  
XX WO200171042-A2.  
PN 27-SEP-2001.  
PD 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
PF 23-MAR-2001; 2000US-0191637P.  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
PA Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX DR  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 25372; 21pp + Sequence listing; English.  
PS The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (AB857737-  
CC AB872072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1601 BP; 344 A; 477 C; 441 G; 339 T; 0 U; 0 Other;  
SQ  
Alignment Scores:  
Pred. No.: 1.51e-64 Length: 1601  
Score: 714.50 Matches: 156  
Percent Similarity: 58.22% Conservative: 60  
Best Local Similarity: 42.05% Mismatches: 140  
Query Match: 33.62% Indels: 15  
DB: 4 Gaps: 6  
US-10-649-273-2 (1-414) x ABL24633 (1-1601)  
Qy 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValaValaAspGluThrGly 58  
Db 79 GTCCTGGGCAATCGAGACTCTCTGCGAGCAACGGGCAATCGCATCTGTGAGACACCGGGC 138  
59 AsnValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValaValaAspGluThrGly 78  
139 CGAGTATTCGCAATGCTGTGAGTCCGACAGAGAGTTCACACCGCTATGAGGCAATT 198

Qy 79 ValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAla 98  
Db 199 ATTCGCGCCAGAGGCCCGAGGACCTTACCGCGCGGATGAGTCCGCTACAGCGGTGC 258  
Qy 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLyPro 118  
Db 259 ATGAGAGCGCGCAATTTGAAGCGGAGCAACTGACGCCCATTCGCGGTGACAGCGGTCCC 318  
Qy 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeu 138  
Db 319 GAGCTGCCCTGATGTGCTGTGGGCGTGCCTTGTGACAGGACCTGCGCCCTCGCTG 378  
Qy 139 LyLyAspProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeuThr 158  
Db 379 CAGAGCCCTGATGTGCTGTGGGCGTGCCTTGTGACAGGACCTGCGCCCTCGCTG 438  
Qy 159 Asn-----LyValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeu 176  
Db 439 CATCCGAGACAGATCGGCTACCCCTTCTGCTGCTGCGCAGCGGCGGCACTGTGAG 498  
Qy 177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLySerLeuAspIleAla 196  
Db 499 TTGTGTGTGCTTACGCGCCCGGTGCGCTTACGCTTGTGCGCAACACTGATGATGCG 558  
Qy 197 ProGlyAspMetLeuAspLyValAlaArgArgLeuSerLeuIleLyHisProGluCys 216  
Db 559 CCCGCGAGGCTTGTGACAGATCGGTCCGACCTACGCTGCACTTCGCGGAGTAC 618  
Qy 217 SerThrMetSerGlyGlyLyAlaIleGlnHisLeuAlaLyS--GlnGlyAsnArgPhe 235  
Db 619 CGCTGTGTGAGACGAGAGGAGCGGCAATTGAGCATCGCGCCAGCTGCGAGCATCGCG 678  
Qy 236 HisPheAspIleLyPheProLeuHisHisAlaLyAsnCysAspPheSerPheThrGly 255  
Db 679 GCCTACGAGTTTCGCTGCGCTGCGCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 738  
Qy 256 LeuGlnHisValThrAspLyIleIleMetLySlyGlnLySerGlnGluGlyIleGluLyS 275  
Db 739 ATCAAGAGACACTCTTCCGCGCATTCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 798  
Qy 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295  
Db 799 GACGAGGTATACGAGCACTACGCGGACCTTCTGCGCGGCTGCTGCGCTGTGCACCGG 858  
Qy 296 HisLeuValLyValArgThrHisArgAlaIleLeuPheCys-----LySgluArgAspLeu 313  
Db 859 CACTGATGTCACCGGACGCGGAGCAATTTGATGCTGCTGCTGCGGACAGGCGAGCTC 918  
Qy 314 LeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrtIle 333  
Db 919 TTCGGTGAACCCCGCCCACTGCTGATGTCGCGGCTGTGTGCAACAACATGCCATTA 978  
Qy 334 ArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro 353  
Db 979 TACGCCAATATGAAATCTGCGCGCGGAGTATGCTGACAGAGCTTCGCTCATGGAAG 1038  
Qy 354 ArgLeuCysThrAspAsnGlyIleMetIleAlaArgAsnGlyIleGluArgLeuArgAla 373  
Db 1039 CGGTACTGCTGCGAACAACCGGCTCATGATGCTGCGACGCGGCTTATGAGCACTG----- 1092  
Qy 374 GlyLeuGlyIleLeuHisAspIleGlu---GlyIleArgTyrtGluProLyCysProLeu 392  
Db 1093 -----CTGCAAGATTAAGAGGCGCAGACGCGCTACGATGATACGAC----- 1131  
Qy 393 GlyValAspIleSerLyGluValGlyGluAla 403  
Db 1132 AGCATTAATTCAGAGGACCGCGGATTCGCG 1164  
RESULT 14  
AAH1510/C  
ID AAH15110 standard; cDNA; 1385 BP.  
XX AAH15110;  
AC AAH15110;

XX 26-JUN-2001 (first entry)  
 XX Human cDNA sequence SEQ ID NO:13144.  
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX Homo sapiens.  
 OS EPI074617-A2.  
 PN 07-FEB-2001.  
 PD 28-JUL-2000; 2000EP-00116126.  
 XX 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX (HELT-) HELIX RES INST.  
 PA Oka T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 DR Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 XX Claim 8; SEQ ID NO 13144; 2537bp + Sequence Listing; English.  
 PS The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dt primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification; where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX SQ Sequence 1385 BP; 439 A; 237 C; 256 G; 453 T; 0 U; 0 Other;  
 XX Alignment Scores:  
 Pred. No.: 3.6e-63 Length: 1385  
 Score: 700.50 Matches: 171  
 Percent Similarity: 38.31% Conservative: 1  
 Best Local Similarity: 38.08% Mismatches: 1  
 Query Match: 32.96% Indels: 277  
 DB: 4 Gaps: 1  
 US-10-649-273-2 (1-414) x AAH15110 (1-1385)  
 QY 74 LysThrGlyGlyLeuValProPheAlaIleGlnGlnLeuHisArgGluAsnIleGlnArg 93  
 DB 1346 AGAAGCAGGTGGATTTCTTCTCCAGCAGCTCAACAGCTTCAAGAAATATTTCAGCA 1287

QY 94 IleValGlnGluAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla 113  
 DB 1286 ATAGTACAGAAAGCTCTTCTCCAGTGAAGTCTCTCCAGAGACCTCTCAGAAATTGCA 1227  
 QY 114 ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln 133  
 DB 1226 ACTACCATTAACCAAGAGACTCTTTAAGCTGGAGAGTGGCTTATCATTTACTTACACAG 1167  
 QY 134 LeuValGlyGlnLeuLysLysProPheIleProIleHisIleMetClnAlaHisAlaLeu 153  
 DB 1166 CTGGTAGAGCAGCTTAACCAAGCATTCTCCATTCATCATATGAGGCTCATGCACTT 1107  
 QY 154 ThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGly 173  
 DB 1106 ACTATTAGGTTCACCAATTAAGTAGAATTTCTTTTATGTTCTTTGATTTCTGAGGT 1047  
 QY 174 HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193  
 DB 1046 CACTGCTGTTGGCATTAAGTTCAGAGAGTTTCAAGATTTCTGCTTCTTGAAAGTCTTTG 987  
 QY 194 AspIleAlaProGlyAspMetLeuAspLysVal----- 204  
 DB 986 GACATATGACACCGGTGACATGCTTGCACAGGT-AAATTAAGAAATTAATTTCTCCATTTCTT 928  
 QY 204 ----- 204  
 DB 927 TTTGTTAGTTGTCATTCAATTTCAATGATGATGATGATGATGATGATGATGATGATGAT 868  
 QY 204 ----- 204  
 DB 867 TATTTCTGATTTTATCTTATGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 808  
 QY 204 ----- 204  
 DB 807 AAGAGTAGACACAATTTATTAATTTCTTACCTTTCTTAATTAATGTAAGAGTTTCAT 748  
 QY 204 ----- 204  
 DB 747 ATCTGTACATAAAGGCTGAATATGTTGATGATGATGATGATGATGATGATGATGATGAT 668  
 QY 204 ----- 204  
 DB 687 ATGTGAAGAAGCTGCTTCTTAACATTAATGATGATGATGATGATGATGATGATGATGATGAT 628  
 QY 204 ----- 204  
 DB 627 ATATGATTATACATTAAGACATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 568  
 QY 204 ----- 204  
 DB 567 TACACCAAGACAGGCTCCCGGACCCCTGTTGTTAGAAATCTACAGAGGCTACT 508  
 QY 204 ----- 204  
 DB 507 GCCATATATGAAAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 448  
 QY 204 ----- 204  
 DB 447 TAGAAGTATAGACAAGTTCTTATTAATGAGCTTCATCATTAAGCAGTATATGCAACT 388  
 QY 204 ----- 204  
 DB 387 TCAGACCATTTTCCACCAATTAAGAGCAAAACATTAAGCAGGCGAGTATGAGCTCT 328  
 QY 204 ----- 204  
 DB 327 TATTGTTGGGTGATCATTAAGAGAGGTTGTCTGCTTACCTGAATATCAGCTATAGTC 268  
 QY 204 ----- 204  
 DB 267 TATATTTGCAAGATATAGATGATTTTATTCATTCAGGAGTTTGTGTTTGTAGTAAT 208



Qy 375 uGlyIleLeuHisAspIleGlu---GlyIleArgTyrGluProLysCysProLeuGlyVa 394  
Db 1509 -----CTGCAGGATTAAGAGGCGCAGCAGCGGCTACGACTACGAC-----AGCAT 1466  
Qy 394 lAspIleSerLysGluValGlyGluAla 403  
Db 1465 TGATATCCAGGCGCAGCGCGGATTGCGC 1438

Search completed: February 16, 2005, 13:58:31  
Job time : 834.4 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:19:37 ; Search time 246.15 Seconds  
(without alignments)  
2752.056 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 2125  
Sequence: 1 MLILTKTAGVFPKSKRKVY.....DISKEVGEASIKVPQIKMEI 414

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues  
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=x1h  
-Q=/cg2n\_1/USPRO.spool/US10649273/runat\_14022005\_114704\_16425/app.query.fasta\_1.1429  
-DB=Issued\_Patents\_NA -OFMT=fastap -SUFFIX=ini -MINMATCH=0.1 -LOOFC=0  
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=p2n -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=p2n -NOEM=ext -HEAPSIZE=500 -MUTLEN=0 -MAXLEN=200000000  
-USBR=US10649273 -CGEN\_1\_1\_292 -runat\_14022005\_114704\_16425 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA:\*  
1: /cg2n\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cg2n\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cg2n\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cg2n\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cg2n\_6/ptodata/1/ina/PTUS.COMB.seq:\*  
6: /cg2n\_6/ptodata/1/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2125	100.0	2197	4 US-10-067-443-1	Sequence 1, Appli
2	2090.5	98.4	1387	4 US-10-067-443-21	Sequence 21, Appl
3	1747	82.2	1416	4 US-09-774-528-177	Sequence 177, App
4	1385	65.2	1526	4 US-10-067-443-23	Sequence 23, Appl
5	1186.5	55.8	14364	4 US-10-067-443-20	Sequence 20, Appl
6	502	23.6	94750	4 US-09-596-002-38	Sequence 38, Appl
7	501	23.6	1053	4 US-09-540-236-806	Sequence 806, App
8	499.5	23.5	1206	4 US-09-252-991A-801	Sequence 801, App
9	492.5	22.2	1059	4 US-09-253-951A-584	Sequence 884, App
10	482.5	22.7	1074	4 US-09-543-681A-2341	Sequence 2341, Ap
11	482	22.7	996	4 US-09-902-540-6612	Sequence 6612, Ap
12	482	22.7	2582	4 US-09-902-540-503	Sequence 503, App

13	479	22.5	1315	1 US-08-087-797-1	Sequence 1, Appli
C 14	475.5	22.4	1830121	4 US-09-557-884-1	Sequence 1, Appli
C 15	475.5	22.4	1830121	4 US-09-643-990A-1	Sequence 2050, Ap
16	474.5	22.3	1032	4 US-09-489-039A-2050	Sequence 2955, Ap
17	472	22.2	1092	4 US-09-107-532A-2955	Sequence 5, Appli
18	461	21.7	1008	3 US-08-987-121A-5	Sequence 51, Appli
19	461	21.7	1011	3 US-08-987-121A-3	Sequence 51, Appli
20	457	21.5	1006	3 US-08-961-083-51	Sequence 51, Appli
21	457	21.5	1006	4 US-09-536-784-51	Sequence 214, App
C 22	457	21.5	10974	3 US-08-961-527-214	Sequence 1, Appli
23	455	21.4	1011	3 US-09-066-512-1	Sequence 1551, Ap
24	454	21.4	1101	4 US-09-134-000C-1551	Sequence 2196, Ap
25	453	21.3	1011	4 US-09-583-110-2196	Sequence 1618, Ap
26	448	21.1	1011	4 US-09-107-433-1618	Sequence 1072, Ap
27	443	20.8	1107	3 US-09-134-001C-1072	Sequence 1, Appli
28	433	20.4	1026	3 US-09-149-624-1	Sequence 102, App
C 29	424	20.0	15249	4 US-08-956-112E-102	Sequence 102, App
C 30	424	20.0	15249	4 US-08-781-986A-102	Sequence 1, Appli
31	406.5	19.1	640681	4 US-09-790-988-1	Sequence 1, Appli
32	406.5	19.1	1230025	4 US-09-198-452A-1	Sequence 147, App
33	406.5	19.1	1230230	4 US-09-438-185A-1	Sequence 794, App
34	404.5	19.0	1155	4 US-09-602-777R-147	Sequence 2, Appli
C 35	404	19.0	3064	3 US-09-221-017B-794	Sequence 1, Appli
C 36	379.5	17.9	4403765	3 US-09-103-840A-2	Sequence 483, App
C 37	379.5	17.9	4411529	3 US-09-103-840A-1	Sequence 19, Appl
38	369.5	17.4	645	4 US-09-328-352-483	Sequence 1, Appli
39	362	17.0	876	4 US-09-724-623-19	Sequence 131, App
40	357	16.8	580073	4 US-08-545-528D-1	Sequence 3985, App
C 41	333.5	15.7	42325	4 US-08-511-731A-131	Sequence 727, App
C 42	327.5	15.4	3993	4 US-09-710-279-3985	Sequence 130, App
C 43	319	15.0	822	4 US-09-710-279-727	Sequence 1, Appli
C 44	317	14.9	36941	4 US-08-311-731A-130	Sequence 1, Appli
45	313	14.7	1664976	4 US-08-916-421B-1	

ALIGNMENTS

RESULT 1  
US-10-067-443-1  
; Sequence 1, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; TITLE OF INVENTION: SPINAL CORD, MP-1  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 2197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (231)..(1472)  
; US-10-067-443-1  
Alignment Scores:  
Pred. No.: 5.72e-261 Length: 2197  
Score: 2125.00 Matches: 414  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
US-10-649-273-2 (1-414) x US-10-067-443-1 (1-2197)

QY 1 MetLeu1leuThrIleuThrAlaGlyValPhePheIysProSerIleuArgIleValTyr 20  
 Db ATGTTAATCTTGAATAAGACTGACGAGGATTTTTTAAACATCAAAAAGAAATTTAT 290  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40  
 Db GAATTTTAAAGAACTTTAAATTTTCATCTCTGAAACACTATTTCTTCATTAATAATAGTATTTG 350  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyIleValPro 60  
 Db GGAATTTAAATAGTATGTGATGATACAGACAGCTGTGTGGTGAATGAACCTGGAATATGTG 410  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisIleuIysThrGlyIleValPro 80  
 Db TTGGGAGAAGCAATACATTTCCCAAACTGAACTTAAACAGAGTGGAGTTTCTTCT 470  
 QY 81 ProAlaIleGlnIleuHisArgGluAsnIleGlnArgIleValGlnIleuIleuSer 100  
 Db CCAGCACTCAACAGCTTCAACAGAAAATTTCAACCAATATGTACAAAGAAAGCTCTTCT 530  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIleuProGlyLeu 120  
 Db 531 GCCAGTGAAGCTCTCCAGAGTACCTTCAGCAATGTGCACTACCATTAACACAGACTT 590  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleuIysIys 140  
 Db 591 GCTTTAAGCTCGGAGTGGCTTTCATTTAGCTTACGCTGTGTAGGACAGTTAAAAAG 650  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160  
 Db 651 CCATTTCATTCCTCATTCATATGAGAGGCTCATGCACTTACATTAAGTTAGCCAAATAA 710  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaVal 180  
 Db 711 GTAAATTTCTCTTTTAAATTTCTTTGATTTCTGAGGTCACTGTGTGGCATTAGTT 770  
 QY 181 GlnGlyValSerAspPheLeuLeuGlyIysSerIleAsnIleAlaProGlyAspMet 200  
 Db 771 CAAGGAGTTTCAGATTTCTCTCTTCTGAAAGTCTTTGGACATAGCCACAGGTACATG 830  
 QY 201 LeuAspIleValAlaArgArgLeuSerLeuIleIysHisProGluCysSerThrMetSer 220  
 Db 831 CTTCACAGAGTGGCAAGAGACTTTCTTTAATAAAACATCCAGAGTGTCCACCATGTAGT 890  
 QY 221 GlyGlyIysAlaIleGluHisLeuAlaIysGlnGlyAsnArgPheHisPheAspIleIys 240  
 Db 891 GGTGGGAAGCCATAGAAACATTTTGCCCAACAGGAAATAGATTTTCATTTGACATCAAA 950  
 QY 241 ProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGlyLeuGlnHisAlaThr 260  
 Db 951 CTCTCCCTTGCACTACAGCTAAATAATGTGATTTTCTTTTACCTGAGCTTCAACACGTTACT 1010  
 QY 261 AspIysIleIleMetIysGlyIleGluGluGlyIleGluIysGlyGlnIleuSer 280  
 Db 1011 GATTAATAATTAATGAAGAAAAGAAAAGAGAGATTTGAGAAAGGGCAATCTGTCT 1070  
 QY 281 SerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisIleuValIysArg 300  
 Db 1071 TCAGCAGACAGCATGCTGCGCACAGTACAGCACCAATGGCATGTCATCTGTGAAAAGA 1130  
 QY 301 ThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuProGlnAsnAsnAlaVal 320  
 Db 1131 ACACATGGGGCTATCTGTTTGTGTAAGCAGAGAGACTTTGTAACCTCAAAATTAATACATA 1190  
 QY 321 LeuValAlaSerGlyIleValAlaSerAsnPheTyrIleArgArgAlaIleuGluIleu 340  
 Db 1191 CTGGTTGCATCTGCTGTGTGTGCAAGTACCTTCATATCCGACAGAGCTCTGGAATTTTA 1250  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaArgLeuCysThrAspAsnGly 360  
 Db 1251 ACAAACCAACACAGTGCATTTGTGTGTCTCTCCACAGCTATGACATGATATAGGC 1310  
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleIleuHisAsp 380

Db 1311 ATTATGATTCAGTGAATAGTATGAAAGACTACAGCTGCGCTGGCATTTTACATGAC 1370  
 QY 381 IleGluGlyIleArgTyrGluProIysCysProLeuGlyValAspIleSerIysGlyVal 400  
 Db 1371 ATAGAGGAGCATCGCTATAGAACCAAAATGTCTTGGAGTGAACATATCAAAAAGAGTT 1430  
 QY 401 GlyIleAlaSerIleIysValProGlnLeuIysMetGluIle 414  
 Db 1431 GGAGAACTTCCATTAAGTACCAATTAATAATGAGATA 1472  
 RESULT 2  
 US-10-067-443-21  
 ; Sequence 21, Application US/10067443  
 ; Patent No. 6642041  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bristol-Myers Squibb Company  
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
 ; TITLE OF INVENTION: SPINAL CORD, MF-1  
 ; FILE REFERENCE: D0073 NP  
 ; CURRENT APPLICATION NUMBER: US/10/067,443  
 ; CURRENT FILING DATE: 2002-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/266,518  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/282,814  
 ; PRIOR FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 21  
 ; LENGTH: 1387  
 ; TYPE: DNA  
 ; ORGANISM: homo sapiens  
 US-10-067-443-21  
 Alignment Scores:  
 Pred. No.: 6,62e-257 Length: 1387  
 Score: 2090.50 Matches: 412  
 Percent Similarity: 93.85% Conservative: 0  
 Best Local Similarity: 93.85% Mismatches: 2  
 Query Match: 98.38% Indels: 25  
 DB: Gaps: 1  
 US-10-649-273-2 (1-414) x US-10-067-443-21 (1-1387)  
 QY 1 MetLeu1leuThrIleuThrAlaGlyValPhePheIysProSerIleuArgIleValTyr 20  
 Db 24 ATGTTAATCTTGAATAAGACTGACGAGGATTTTTTAAACATCAAAAAGAAATTTAT 83  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40  
 Db 84 GAATTTTAAAGAACTTTAAATTTTCATCTCTGAAACACTATTTCTTCATTAATAATAGTATTTG 143  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyIleValPro 60  
 Db 144 GGAATTTAAAGAACTTTAAATTTTCATCTCTGAAACACTATTTCTTCATTAATAATAGTATTTG 203  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisIleuIysThrGlyIleValPro 80  
 Db 204 TTGGGAGAAGCAATACATTTCCCAAACTGAAATTTCAATTTAAACAGTGGGATTTGCTCT 263  
 QY 81 ProAlaIleGlnIleuHisArgGluAsnIleGlnArgIleValGlnIleuIleuSer 100  
 Db 264 CCAGAGCTCAACAGCTTCAACAGAAAATTAATTAACGAATATGTATCAAGAAAGCTTTTCT 323  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIysProGlyLeu 120  
 Db 324 GCCAGTGAAGCTTCTCCAAAGTACCTTCAAGATTTGCACTACCATTAACCAAGACTT 383  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyIleuIysIys 140  
 Db 384 GCTTTAAGCTCGGAGTGGCTTTCATTTAGCTTACGCTGTGTAGGACAGTTAAAAAG 443  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnIys 160







Best Local Similarity: 32.20%		Mismatches: 2	
Query Match: 55.84%		Indels: 657	
DB: 4		Gaps: 4	
US-10-649-273-2 (1-414) x US-10-067-443-20 (1-14364)			
QY	74	LysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg	93
DB	10623	AGAACAGGTGGAGATTGTTCCCTCCAGACGCTCAACAGCTTCCACAGAAATATTCACAGA	10682
QY	94	IleValGlnGlnIleLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla	113
DB	10683	ATAGTACAGAGAGCTTTCTGCGACAGAGCTCTCCAAAGTAGACCTCCAGCAATTGCA	10742
QY	114	ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln	133
DB	10743	ACTACCAATTAACAGAGCTTTGTAAGCTGGAGAGGGCTTATTCATTAGCTTACAG	10802
QY	134	LeuValGlyGlnLeuLysLysPheProPheIleProIleHisIleMetGluAlaHisAlaLeu	153
DB	10803	CTGTAGAGACAGTTAAATAAACCATTCATTCCTCATCATATGAGAGCTCATGCACTT	10862
QY	154	ThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGly	173
DB	10863	ACTATTAGTTGACCAATTAAGTAGAATTTCTTTTATTTAGTTCTTTGATTTCTGAGGT	10922
QY	174	HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu	193
DB	10923	CACGTCTGTTGGCAATTAGTTCAAGAGATTTCAGATTTTCTGCTTCATGAAAGTCTTGG	10982
QY	194	AspIleAlaProGlyAspMetLeuAspLysVal	204
DB	10983	GACATGACCAAGGTGACATGCTTGACAGAGGT -AATTAGAAATTAATTTCTCCATCTTT	11041
QY	204	-----	204
DB	11042	TTTGTATATGTTGTCATTTTCACTAAGTAGCAATAGATGTCATCACCATTCACCTAAA	11101
QY	204	-----	204
DB	11102	TATTTCTGAATTTTATCTTAGTAAATCTGAAAAAATTCACATATGTCAGAAAAATAGA	11161
QY	204	-----	204
DB	11162	AAGAGTAGTACACAAATTTTATTAATTTCTAGCCTTTCTTAATTAATGAGTAGAGGTTGAT	11221
QY	204	-----	204
DB	11222	ATCTGTACATTAAGGCTGAATTAAGTTTGACATACAGTTATGTAATTTGCCAAATTAATGT	11281
QY	204	-----	204
DB	11282	ATGTGAAGAAGCGTCTTCGTAACATAATGCAATGCAAAAAAGTTAAATAGAGATAT	11341
QY	204	-----	204
DB	11342	ATATAGATTAACTATAGACATTAAGAATGCAATGCAAGAAATTAATCACAATTAAT	11401
QY	204	-----	204
DB	11402	TACACCAACAGACAGGTCCTCCCAAGGCTTTGTTTATGAAATACATACAGAGGCTACT	11461
QY	204	-----	204
DB	11462	GCCATATATAGAAAACAAACAAACAAACAAACAAACATGCTTCCACAGTGAATATA	11521
QY	204	-----	204
DB	11522	TAGAAATATAGACAGATTTCTTATTAATGACGTTACATTAAGCAGTTATTTGCACT	11581
QY	204	-----	204
DB	11582	TCAGCCCATTTTTCACCAATATAGAGACAAACATATAGACAGGGGAGTGAATGGCTCT	11641

QY	204	-----	204
DB	11642	TATTTGGGTCATCATTAAGAACAGGGTGTCTGCTTAAGTAATTCAGCTATATGTC	11701
QY	204	-----	204
DB	11702	TATATTGGCAAGATATAGCATGTTTATTCATTCAGGGGTTTTTTGTTTGTATGTAAT	11761
QY	204	-----	204
DB	11762	TTTCAATTAATTTCTTTGTCATCTTTTCGTTTCACAGTATTAATTAATTAAGCTTAAA	11821
QY	205	-----	205
DB	11822	AATATGTTTCTTTGATAGTGCCACAGAGCTTTCTTTAATTAATTAATTCAGAGTCTCC	11881
QY	218	ThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPhe	237
DB	11882	ACCATGAGTGGTGGGAAACCATTAAGACATTTGGCCAAACAGGAATATGATTTCAATTTT	11941
QY	238	AspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGln	257
DB	11942	GACATCAAACTCCCTTGATCATATGCTTAAATTTGATTTTCTTTACTGGAATTCAA	12001
QY	258	HisValThrAspLysIleIleMetLysLysGlyLysGlnGlu	271
DB	12002	CACGTTACTGATTAATTAATTAATTAATTAATTAATTAATTAATTTCTAATTAATGCT	12061
QY	272	-----	272
DB	12062	AAAGTTGAACAGATTAATATTCCTGATGTCGCTTAAATAGCTGCTCATTTCTGACAG	12121
QY	272	YlleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThr	292
DB	12122	TATGAGAAAGGGCAAAATCCTGCTTCACAGACAGACATTTGCTGCCACAGTACAGACAC	12181
QY	292	MetAlaCysHisLeuValLysArgThrHisArgAlaAlaLeuPheCysLysGlnArgAs	312
DB	12182	AATGGCATGTCATCTTGTGTAAGAAACACATGGGCTATTTGTTGTAAGCAGAGAGA	12241
QY	312	PleuLeuProGlnAsnAsnAlaValLeu	321
DB	12242	CTTGTTACTCAAAATTAATAGCATGCTGTAAGTTTATCTCATTTTATATGTAATGTTA	12301
QY	321	-----	321
DB	12302	CACTTGCAATATGTTACTTTTTCACAGACCTTGACCTTGTTGTTAGAGTACAGAT	12361
QY	321	-----	321
DB	12362	CTTTATGCTTATGCTAGCCCTGACAGATGAATTAATGACAGATAGAAAGCTAAACAG	12421
QY	321	-----	321
DB	12422	CCATTTCTTGTAAGTTGTAGCTTTATGAGCAGCTGTATAGCTTCTATGACACATA	12481
QY	321	-----	321
DB	12482	AGCTTAATTTTGATCTTCTGTTGATTTTAAAGAGGGCTTACATTAAGAAAGTAAAT	12541
QY	321	-----	321
DB	12542	GCAATGACTGATATCATATTTTATGAAAAATAGGTGATTTCTTCATCTTTGATGAA	12601
QY	321	-----	321
DB	12602	ATCCCTTGTTGTTGTTTATTAATTAATGACAGTCAATTAATAGCAGTGGAGGTGTAT	12661
QY	321	-----	321
DB	12662	TCCAACTTGTGACATATATGTTGATTAAGTTTGTATATCATATATTTGTACAGCC	12721

```

QY      321 ----- 321
Db      12722 AAAATCCCTTAATGCTTAAAGCCTGACAAACATCCTGTTAACTGATCTTAAA 12781
QY      321 ----- 321
Db      12782 CTTTATTCATTTAAAAATTAATACTAAAGTGGAAAAATGTTAAATGTAATTCAT 12841
QY      321 ----- 321
Db      12842 AGATGGAATTTTACATGATATACAAAGAAATAATTTTTTTCAGAGTTATGTAAATGCA 12901
QY      321 ----- 321
Db      12902 CAAAAATAATAAAATTTTCAGGCTCTAAATAATGCTACTATGATGAAATTATATTAATA 12961
QY      321 ----- 321
Db      12962 AATATTAGATGAAGGTGGAGAAAATATACAAAAATGCTAGTAATGTTGTATGCTA 13021
QY      321 ----- 321
Db      13022 TTAGAAATTAAGTAATTTTTTTCTTCCAAATTTTATACATGATATGTCATCTGCC 13081
QY      321 ----- 321
Db      13082 CATTAACCATCTCAAAAATGGAGATGTTATATTTGTTAAATGCTGATATTTTTCTCCAG 13141
QY      321 ----- 321
Db      13142 TTTAATTAGACGCTTGTCATATCATATATGATGATTTATTTGTTTCTCAATTCCT 13201
QY      322 ----ValAlaSerGlyValAlaSerAsnPhenylleuArgAlaLeuGluIle 340
Db      13202 TCAGGTTCATCTGGTGGTGTGCGCAAGTAATCTTATATCCGACAGCTCTGGAAATTTT 13261
QY      340 uhhraaenalaethglnctsrthleuleucysproprobraleucysrthraasngl 360
Db      13262 AACAAACGCAACACAGTGCATTTGTTGTCTCTCCACAGCATATACATGATTAATGG 13321
QY      360 yllemecllealatr----- 365
Db      13322 CATTAATGATTCATGTAAGCCACAGATATAGGTGCTTCACTCATTAATGTAATAT 13381
QY      365 ----- 365
Db      13382 TAATGCCATTTATCATACTAAGCCTTCTTCTTCAGATCTTGGAGCTATTGATTTAT 13441
QY      366 -----AsnGlyIleGluAlaGluArgAlaGlyLeuGlyIleuH 379
Db      13442 TTTAATGCTCTTATTTAGGAATGGTATTTGAAAGCTAAGCTGCTGGCATTTTAC 13501
QY      379 iAspIleGluGlyIleArgTyrGluProlys 389
Db      13502 ATGACATAGAGGCAATCCGCTATGACCAAG 13533

RESULT 6
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Paterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38

```

	LENGTH: 94750	
;	TYPE: DNA	
;	ORGANISM: M. catarrhalis	
;	FEATURE:	
;	NAME/KEY: misc feature	
;	OTHER INFORMATION: Incyte template ID No. 6632636 38	
;	PUBLICATION INFORMATION:	
US-09-596-002-38		
Alignment Scores:		
Pred. NO.:	2.6e-50	Length: 94750
Score:	502.00	Matches: 129
Percent Similarity:	50.81%	Conservative: 59
Best Local Similarity:	34.86%	Mismatches: 149
Query Match:	23.62%	Indels: 34
DB:	4	Gaps: 9
US-10-649-273-2 (1-414) x US-09-596-002-38 (1-94750)		
QY	39	ValleuGlYlleglUthrSerCySaPAPRThraAlaAlaValaspGlUthr---
Db	23873	GTATTGGATTGGAGACATCTTGATGTGAACAGGGCTTGCCTCATGATGTATGACGATG
QY	58	-----GlyAenValleuGlYlAlaAlaIleHisSerGlnThrGlValHisleu
Db	23933	AATGGCGGTGCGGGCGGCGTGTCTGTCTCAGGTTTGTATGACCAATTAATCAATGCC
QY	74	LyethrGlYlglYlAlaProProAlaAlaGlnGlnLeuHisargGluAsnIleglnArg
Db	23993	ACCATATGGCGGTGTGCGTGCCTTGAAGCTTGCACAGTCAGACACACTTGTGAAGCTTGCCG
QY	94	IleValGlnGluAlaLeuSerSerAlaSerGlyValSerProSeraspLeuSerAlaIleAla
Db	24053	TTATTATTAAGCTGTGTGATCAGGCAAAATATACCAAAATCCGACATGTATGCGGTGGCG
QY	114	ThrThrIleLysProGluLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln
Db	24113	TATACCAAAAGGGCCAGGGCTGATTTGGGCGCTTATATGACAGGGGCATTTATTTGGCGGACG
QY	134	LeuValGlyGlnLeuLysProPheIleProIleHisIleMetGluAlaHisAlaLeu
Db	24173	CTGGCCGATATGGGCTGGCGGCGCCAGCGGCTTGGGGGTGCATCATATGAGAGGACATCTGTTA
QY	154	ThrIleArgLeuThr-----AsnLysValGluPheProPheLeuValLeuLeuIleSer
Db	24233	GCAACGGTATTTGGCCAGATGATGACCCCCCAAGCTTTCCTGTTGTGCGCTGCGGTGCG
QY	172	GlyGlyHisCysLeuLeuAlaLeuAlaGlnGlyValSerAspPheLeuLeuLeuGlyLys
Db	24293	GGCGGTATACCAATGCTGTCGTGCCGTGCCAGATGGTGGCGGTGATGATATATTGGGCGAG
QY	192	SerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgSerLeuIle
Db	24353	TCATTCATATATGCGGCGGTGAGTGAAGCTTGTATATTAACGCCAATAATGCTCAACATG---
QY	212	LysHisProGluCysSerThrMetSerGlyGlyLysAlaIleglnHisLeuAlaLysGln
Db	24410	CCCATATCT-----GGTGGCCCAAAATATGCAAAATATTAGCCAAAAAC
QY	232	GlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLysAsnCySaPhe
Db	24452	GGCAACCCACACGCTTATGAGTGCACCAAGACCCATCAGCAT---AAAGGCTGTGATTTT
QY	252	SerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysGlyGluGlu
Db	24509	TGCTTATGAGGACATGAACACCGGCATTCATATCTCATTAAGACACACACCAACCGCCCA
QY	272	GlyYlleglUysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHis
Db	24569	AGGACGCC-----GCCACACGAGACATGCGCCCAAGCTTGAAGTAT
QY	292	ThrMetAlaCysHisLeuValLysValArgArgHisArgAlaIleLeuPheCysLysGlnArg

```

Db      24614 GCGGTGTGATCTTTGGTCAAAAATGCAACCAAGCACTACAGATGACAGCATTCGC 24673
Qy      312 AspleuleuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAspHe 331
Db      24674 CAG-----CTGGTGTGCGCAGGGGGCGTCTCTGCAATTCAG 24709
Qy      332 TyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysPro 351
Db      24710 ATGCTAGCCCGCACCTTCAAGCAGAGCGCTCCGCAATGATGCGTGGTATCTATGCGC 24769
Qy      352 ProProAlaGlyCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeu 371
Db      24770 CCGACCGAGCATATGACGGAATATGATGCGATGCTGCTGCTTTGTGCGGCTC 24829
Qy      372 ArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyTrpGluProLysCysPro 391
Db      24830 AGCTGTGACAGCTGC-----GATACCTTGGC-GGTTGCGCTGTATTC-----CCG 24873
Qy      392 LeuGlyValAspIleSerLysGluValGly 401
Db      24874 ATGGGATATGACGACGCTTGGCGTATCGGC 24903

RESULT 7
US-09-540-236-806
; Sequence 806, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 806
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: M.catarhalis
; US-09-540-236-806

Alignment Scores:
Pred. No.:      1,23e-53      Length:      1053
Score:          501.00      Matches:      126
Percent Similarity: 49.86%      Conservative: 56
Best Local Similarity: 34.52%      Mismatches: 139
Query Match:      23.58%      Indels:      44
DB:              4          Gaps:      8

US-10-649-273-2 (1-414) x US-09-540-236-806 (1-1053)
Qy      39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThr--- 57
Db      16 GTATTGGGATTGGAGACATCTTGATGAAACAGCGCTGCGCATGATGATGATGACATG 75
Qy      58 -----GlyAsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLys 74
Db      76 AATGGCAGGGGGCGGCGTCTCTCTCAAGTTTATATACCAAAATTATATCATATCCACC 135
Qy      75 ThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIle 94
Db      136 TATGGCGGTGTCGTCCTGACCTTGCAGTCAGACACCATTCGTAAGCTTGTGCCGTTA 195
Qy      95 ValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspSerLeuSerAlaIleAlaThr 114
Db      196 TTTAATAGATGATTGATCAGGCAAAATATACCAAAATCGAATTCGATGCGGTGCGTAT 255
Qy      115 ThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeu 134
Db      256 ACCAAAGGCCCGGGCGCTGATTTGGGGATATGACAGGGGCGCATTTTGGCGCGACGCGT 315
Qy      135 ValGlyGlnLeuLysLysProPheIleProIleHisIleMetGluAlaHisIleAlaLeuThr 154
Db      316 GCGTATGGCGGTGGCGTCCAGCGGTGGGGTGCATCATATGAGAGGACATCTGTACGA 375

```

```

Qy      155 IleArgLeuThr-----AsnLysValGluPheProPheLeuValLeuLeuIleSerGly 172
Db      376 CCGCTATTGGCCAGTATATGACCCCAAGCTTTCCGTTGTGTGCTGCTGTGTGCGGC 435
Qy      173 GlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSer 192
Db      436 GGTATACCATCATGTGTGTGCGTGCATGGTGTGGCGGTATATGACATATTTGGCGAGTCC 495
Qy      193 LeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLys 212
Db      496 ATCGATGATGCGGTGGCGCAATGCTTGTATAAACGGCAAAATGCTCAAACTG---CCC 552
Qy      213 HisProGluCysSerThrMetSerGlyLysValAlaIleGluHisIleuAlaLysGlnGly 232
Db      553 TATCTT-----GGTGTCCAAATATGCAAAATTAATGCAAAACGCG 594
Qy      233 AsnArgPheHisPheAspIleLysProProLeuHisIleAlaLysAsnCysAspPheSer 252
Db      595 AACCCACAGCCCATATGAGCTGCCAAGACCGATGCAGAT---AAAGGCTGATTTTTCG 651
Qy      253 PheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGly 272
Db      652 TTCAGTGGCATGAAAACCCCATTCATTAATCTCATCAAGACACAGCCCAACGCCAAAGC 711
Qy      273 IleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaThrValGlnHisThr 292
Db      712 GACCTT-----GCCACACGACGACATGCGCCGCAAGCTTGTAGTATGCG 756
Qy      293 MetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAsp 312
Db      757 GTGTGTGATATCTTGTGTCAAAATATGACCAAGCATACAGATGACAGCATTTCCGACG 816
Qy      313 LeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaAspAspPheTyr 332
Db      817 -----CTGGTGTGCGCAGGGGCGTCTCTGCAATCAGACG 852
Qy      333 IleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProPro 352
Db      853 CTACCGCCGACCCCTTACCGACGAGCGCTGCGCAATGATGCGGTGCTGACTACGCCCCG 912
Qy      353 ProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArg 372
Db      913 ACCGAGCTATGACCGGATATGCTGCGATGATCGCTTACGCTGCTTTGTGCGGCTAAGC 972
Qy      373 AlaGly-----Leu 375
Db      973 CGTGGGCACTGCGAGTATGCGCGTTCGTATTTCCCGATGGATATGACATGCTT 1032

RESULT 8
US-09-252-991A-801
; Sequence 801, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 801
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

```

US-09-252-991A-801

## Alignment Scores:

Pred. No.: 2,43e-53 Length: 1206  
 Score: 499.50 Matches: 133  
 Percent Similarity: 51.68% Conservative: 67  
 Best Local Similarity: 34.37% Mismatches: 158  
 Query Match: 23.51% Indels: 29  
 Gaps: 4 10

US-10-649-273-2 (1-414) x US-09-252-991A-801 (1-1206)

```

QY 13 LysProSerLysArgLysValTyrGluPheLeuArgSerPheAsnPheHisProGlyThr 32
DB 91 AAAGCTCCAGTCAACAAACCTGCTCGGCGCAAAAGCCGCGATTATCATCGCGGCTT 150
QY 33 LeuPheLeuHisLysIle-----ValLeuGlyTyrLeuThrSerCys 46
DB 151 CGTATTTCGCAAGCTCTAAGGCTCAAGCCATGCGCGTCTGGACTGGAAACGTCTTC 210
QY 47 AspAspThrAlaAlaValValAspGluThrGlyAsnValLeuGlyGluAlaIleHis 66
DB 211 GACGAAACCGGCGCTCGCTTACGACAGCAAGCCGCGCTGCGCGCATTC 270
QY 67 SerGlnThrGluValHisLeuLysThrGlyGlyIleValProAlaIleGlnGlu 86
DB 271 AGTCAGATCGACCTCATCGCTACGCGCGCGCTGCTCCGAGCTGCGCTCGCGCAC 330
QY 87 HisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSerAlaSerGlyValSerPro 106
DB 331 CACGTCAGACCGCATGCTCGCTGATCGCGCCAGGTCGTCGACGATCGCGTCGACGCG 390
QY 107 SerLeuLeuSerAlaIleAlaThrThrIleLysProGlyLeuAlaLeuSerLeuGlyVal 126
DB 391 GCGATATCGACGCGATCGCTATACCGCGCGCTCGCGCTGCTGCTGCTG 450
QY 127 GlyLeuSerPheSerLeuGlnLeuValGlnLeuLysLysProPheIleProIleHis 146
DB 451 GGGGCTTCTTGCTGCCAGGCGATGCGCTTGGCGCGCGCTGCGCGCTGCGCTG 510
QY 147 HisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPhe 165
DB 511 CACATGGAAGGCGACCTGCTGCGCGCGCGATGCTGGAAGACACACCGCGGTTCCGTTCC 570
QY 166 LeuValLeuLeuLysSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAsp 185
DB 571 GTCCGCTTGTGGGCTTCCGCGGCTGACACCCAGTGTGCGCGCTGATCGCGCTG 630
QY 186 PheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAla 205
DB 631 TACCAATTGCTTGGGAATCGGTGACATGCGCGCGGGAAGCCTTGCACAAAGACCGCC 690
QY 206 ArgArgLeuSerLeuLysLeuHisAspProGlyLysSerThrMetSerGlyLysAlaIle 225
DB 691 AAGCTGATCGGCGCTG---GGCTATCC-----GGTGTCCGGAATATC 729
QY 226 GluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProPheLeuHisHis 245
DB 730 GCCCGCTGCGCGAGCGCGGCACTCTGCGCGCTGCTGCTGCTGCGCGCGATCCGAT 789
QY 246 AlaLysAsnGlyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMet 265
DB 790 CGCCCGGCGCTGAGACTTCAAGCTTCAAGCGGCTCAAGACCTTATACCTTGAC--ACTGG 846
QY 266 LysLysGluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaIleAspIle 285
DB 847 CAGCGTTCCTCGAGCGCGCGGACGACGACGACGACG-----ACCGCTGCGCATC 897
QY 286 AlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIle 305
DB 898 GCCCTGCGGCTTCAAGACGCGGCTGTGTCGAGACCTGCTGATCAAGGCGCGCTG 957
QY 306 LeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGly 325

```

```

DB 958 -----AACGACCGCGCTG-----AAGAAC-----CTGGATCGCGGC 993
QY 326 GlyValAlaSerAsnPheThrIleArgAlaIleGluIleLeuThrAsnAlaThrGln 345
DB 994 GGTGTCAAGCCCAACGAGCGGCTGCGCACGCGCTGGAAGATGCTCGCGGAATGAC 1053
QY 346 CysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 365
DB 1054 GGGCAGGTGTTACGCGCGCGCGCGCTTCTGACCGACAAATGCGCGATGATCGCTAC 1113
QY 366 AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArg 385
DB 1114 GCGCGTGCACAGCGCGCTGCTCGCGCG-----CAGCATGACGCGCGCGATCAGC 1164
QY 386 TyrGluProLysCysProLeu 392
DB 1165 GTCCAGCCGCGCTGCGCGATG 1185

```

## RESULT 9

```

US-09-252-991A-884/C
; Sequence 884, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 884
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884

```

Alignment Scores:  
 Pred. No.: 1.52e-52 Length: 1059  
 Score: 492.50 Matches: 128  
 Percent Similarity: 53.24% Conservative: 61  
 Best Local Similarity: 36.06% Mismatches: 143  
 Query Match: 23.18% Indels: 23  
 Gaps: 9

US-10-649-273-2 (1-414) x US-09-252-991A-884 (1-1059)

```

QY 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGly 58
DB 1038 GTGCTGGAGCTGGAACGTCCTGCGCAAGAACCGGCGGTGCCCTTACGACAGCAAGCG 979
QY 59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyIle 78
DB 978 GGCCTGCTGCGCGACCGCTGTTCACTGATCGATCGACTCATCGCTTAAGCGCGGCTC 919
QY 79 ValProProAlaAlaGlnGlnLeuHisAspGluAsnIleGlnArgIleValGlnGluAla 98
DB 918 GTGCCGAGCTGCGCTTCCGCGGACACGTCMAAGCGATGCTGCGCTGATCCGCGAGTG 859
QY 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysPro 118
DB 858 CTCGACAGATCCGGGCTGACGCGCGGATATGACGCGATGCTATATACGCGCGTCCC 799
QY 119 GlyLeuAlaLeuSerLeuGlyValGlyLysSerPheSerLeuGlnLeuValGlnGlnLeu 138
DB 798 GGCCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 739
QY 139 LysLysProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThr 158

```

```

Db 738 GAGCGCCGGCGGCGCGTCGACCAATGAAAGGCACTGCGCCGATGTCGAA 679
Qy 159 Aenlyr---ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeu 177
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 678 GAGCAGCCACCGCGGTTCCCGTTCCGCTTGGTTCCTGCGGCTACACCCGCTTG 619
Qy 178 AlaleuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaPro 197
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 618 GTGCGGAGTGAAGGATACGGCGCTACCAAGTTCCTGGCAATCGGTGACGATGCGGCC 559
Qy 198 GlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSer 217
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 558 GCGCAAGCGCTTCGACAAAGCCGCAAGCTGATCGGCTG---GGCTATATCC----- 511
Qy 218 ThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyLysAsnArgPheHisPhe 237
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 510 -----GGTGGCCCGGAATCGCCCGCTGGGAGAGCGGCACTCTCGCCGCTTC 460
Qy 238 AspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeuGln 257
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 459 GTGTTCCCGCGCGCGATGACCGATCGCCCGCGCTGACATTGACGTTCAAGCGGCTCAAG 400
Qy 258 HisValThrAspLysIleIleMetLysGlyLysGlyGlnGlyIleGluLysGlyGln 277
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 399 ACCTTACCTTGAC---ACCTGCGACGCTGCTCGAGCGCGCAGCAGCAGCAGCAG 343
Qy 278 IleuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeu 297
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 342 -----ACCGCTGCGACATCGCCCTGCGCTTCAGACCGCGGTGTGACACCTTG 292
Qy 298 ValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGln 317
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 291 CTGATCAAGTCCCGCTCGCGCTTG-----AGCAGACCGGCGCTG-----AAGAAC 247
Qy 318 AsnAlaValLeuValAlaSerGlyGlyValAlaLysAsnPheTyrlleArgArgAlaLeu 337
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 246 -----CTGGTATCGCGCGCGGTGTGACGCGCAACAGCGCTGCGAGCGCGCTG 196
Qy 338 GluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThr 357
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 195 GAAAGAAGATGCTCGCGGAATGAGGGGAGGTGTTCTACGCGCGCGCGCTTCTGACAC 136
Qy 358 AspAsnGlyIleMetIleAlaIleArgAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyLe 377
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 135 GACATATGCGCGATATGCTCTACCGCGCTGCGCGCTGCTGCTGCGCGCGC----- 85
Qy 378 LeuHisAspIleGlnGlyIleArgTyrlleArgTyrlleArgTyrlleArgTyrlle 392
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 84 CAGCATATGCGCGCGCGATGACGCTCAAGCGCGCGCTGCGCGCATG 40

```

```

Percent Similarity: 49.58% Conservative: 55
Best Local Similarity: 34.17% Mismatches: 149
Query Match: 22.71% Indels: 31
DB: 4 Gaps: 8
US-10-649-273-2 (1-414) x US-09-543-681A-2341 (1-1074)
Qy 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 58 GTTTTAGGTTTAAACATTTTCCGATGAAACCGGATGCGCAATTACGATGATAAAGCC 117
Qy 59 AsnValLeuGlyLysAlaIleHisSerGlnThrGlnValHisLeuLysThrGlyGlyLe 78
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 118 GGTCTGTAGCGAATCAACTATACGCAAAATAAACTGCAACCGCGCACTATGAGTGT 177
Qy 79 ValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnAla 98
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 178 GTTCCGAGCTTGTCCACGAGTCAATTCGCAAAACAGTGCACATGATCCAAAGCGCC 237
Qy 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysPro 118
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 238 CTTAAAGAAACAATTAACTGCAAAAGATATGATGCGGTGCTTATCTGACAGACCA 297
Qy 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeu 138
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 298 GGGCTGTAGGGGCAATTGCTGCGTGGGGGAGCAGATGCTGCTGCGGCTTGATG 357
Qy 139 LysLysProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThr 158
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 358 GATGTCCTGTATTCACATTCACATATGAAAGGCGCTTATTAAGCGCGCATG 417
Qy 159 AsnLysVal---GluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeu 177
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 418 GAGAAAGCCAGATTTCTTCTGCGGTATCTGTCTCGGGGGGCGCATACCAATTA 477
Qy 178 AlaleuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaPro 197
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 478 ATTAGTGTAAACAGGATTTGGGAAATATACCTGTTAGTACGCTATGATGATGCTGCT 537
Qy 198 GlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSer 217
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 538 GGTGAAGCATTTGATTAACAGCCAAAGCTATTTGGGCTT---GATTATCC----- 585
Qy 218 ThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyLysAsnArgPheHisPhe 237
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 586 -----GGCGGCGCTGTTTATCAAAAATGCAACAACAGGTGTAGAGAGCGTTT 636
Qy 238 AspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGlyLeuGln 257
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 637 GTTTTCTCTGCTCCATGACAGACAGACCGCGACTGACTTTAGTTTCTGCTTAAAA 696
Qy 258 HisValThrAspLysIleIleMetLysGlyLysGlnGlyGlnGlyLysGln 277
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 697 ACCTTGGCGGTAATGATCTATGCTCAAAAGCATGATTCAGAG----- 738
Qy 278 IleuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeu 297
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 739 -----CAAACTCGACAGATATGCGCGCTGTTGAAGATCCGCTAGATATCTTGG 792
Qy 298 ValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGln 317
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 793 GCAATTAATGTCGTGACCA-----TTAGAGCAACA 825
Qy 318 Asn---AlaValLeuValAlaSerGlyGlyValAlaLysAsnPheTyrlleArgArgAla 336
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 826 GCGTTTAAACGCTTAGATGATGCTGGGCGCTGAAGCTCAACGCTTACCGCGCAAA 885
Qy 337 LeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCys 356
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 886 ATGCGAGATATATAGAACAACTCGAGGGGAAGTGTATTATGCTGCGCGCTGATATGT 945
Qy 357 ThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGlnArgLeuArgAlaGlyLeu--- 375
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Alignment Scores: 2.97e-51 Length: 1074  
Pred. No.: 482.50 Matches: 122

Db 946 ACCGATATAGTGCATGATTCCTTTGGCGGAGATGATCCGTTTAAAGTGTACCGAG 1005  
Qy 376 GYILEUHSASPILLEGLUGLYILEARGTYRGUPLROLYCYSPROLEU 392  
Db 1006 GGGGCTTTA-----GGGGTGAACGTGACACACGTTGGCCTTTA 1044  
RESULT 11  
US-09-902-540-6612  
; Sequence 6612, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIORITY FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIORITY FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 6612  
; LENGTH: 996  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-6612  
Alignment Scores:  
Pred. No.: 3,01e-51 Length: 996  
Score: 482.00 Matches: 128  
Percent Similarity: 52.48% Conservative: 52  
Best Local Similarity: 37.32% Mismatches: 131  
Query Match: 22.68% Indels: 32  
Gaps: 9  
DB: 4  
US-10-649-273-2 (1-414) x US-09-902-540-6612 (1-996)  
Qy 38 ILEVALLEUGLYILEGLUTHISERCYASAPSPTRHAIALAVALAVALASPGLUTHR 57  
Db 4 CTCGCTCCTGAGCTCGAAACCTCGTGTATGAGACTGCGCGCTCGTGAGAGACGCGC 63  
Db 58 GLYSANVALLEUGLYILEGLUTHISERGINTHRGUVALHISLEUYSRTHRGU 77  
Qy 64 CGCGCGCGCTGTCGATGTCGTCACGAGTGAGATCAGCGCGCTCGTGAGAGACGCGC 123  
Db 78 ILEVALPROPLAALAGLNGLEUHSARGUASRIEGLARGILEVALGNGLU 97  
Qy 124 GTGTGTGCGGAGCTGCGACCGCAACATCGTCCAGGTGCGCGCTCGTCCACGAG 183  
Db 98 AIALEUSERRALASERGLYVALSERPROSERAPLEUSERRALALEATHRTHRIEY 117  
Qy 184 GCGCTGACGCGGCGCAACAGCGCTCGACGAGTGCATCGCGCTCGACGTCGCGC 243  
Db 118 PROGLYLEUVALLEUSERTLEUGLYVALGYLEUSERTPHERSERLEUGLNUVALG 137  
Qy 244 CCCGAGCTCATCGCGCGCTGCTGTGTGAGAGTGAGTGCGCAAGCGCTTGAGCGCTG 303  
Db 138 LEUYSLEYSRPROPHILEPROLLEHISHISMETGLUALHIALALEUTHRILEARGL 157  
Qy 304 ACGGCGAAGCCCTTGTTGTGGCGCCCAACACCTCGAGAGGCGCATCTGCGCCATCG 363  
Db 158 THRANLYSVAL--GLUPHEPROPHLEUVALLEUULENISERGLYGLYHISCYSL 176  
Qy 364 TTGGAGTGTGCGCGCGAGCGCGCTTCCTTGCGCTCGTTCGCGCGGACACACAGC 423  
Db 177 LEUVALLEUVALGNGLYVALSERAPHELEUVALLEUGLYLYSERLEUASPLLEA 196  
Qy 424 CTCTACGAGGTGACGCGCTTACGCGGACGATCCGCTGTGGGAGAGACGCGCAACG 483  
Db 197 PROGLYAPMETLEUASPLYSVALAALARGLEUSERTLEUILEYHISPROGLUCY 216  
Qy 484 GCGGCGGAGGCAATATGACAGACCGCTGCACTCTCGGCTG---CCGTAATCG----- 534

Qy 217 SERTHMETSERGLYLYLSALAIIEGLUHSLEUALALYSGNGLYASN----- 233  
Db 535 -----GTTGGGAGCCCATCGACAGTGTGGCGAGGAGAACCGGAGGCC 582  
Qy 234 ---ARGPHEHISPPHEASPILEYSRPROLEUHSIALALYASNCYSAPRPHESR 252  
Db 583 ATCCGCTTC-----CCGCGCGCGCTCGCGGCGCAACATTCGACGTCGTC 627  
Qy 253 PHEHTRGLYLEUGLHISVALHTRAPRYSILEIEMETLYLSGLUYSGLUGLU 272  
Db 628 TTCTCCGGGTGAG-----ACGGCGTCTCACCACTGCAAGACGCGC 675  
Qy 273 ILEGLUYSGLYGLILEUUSERSERLAIASPILEALATHRVALGNIHSTR 292  
Db 676 GTCCCGGAGGAGGAGCGCTG-----GGGATTTGTGCGGCTTCCAGAGGCC 726  
Qy 293 METALACYSHISLEUVALYLSARGTHRIASARGALAIIEUPHECYSLYSGINARG 312  
Db 727 GTGGCGGAGCTGTGTGGAAG-----CTGGTGGCGCGCGCGCGCG 771  
Qy 313 LEUUPROGLINASBNALAVALLEUVALALASERGLYVALALASERAPHE 332  
Db 772 TTG-----GGCCACAGCAGTGTGTGTGCGCGCGCGCTCGCACTGCGCG 822  
Qy 333 ILEARGVALALEUGLUILEUTHRANALATHRGINCYSTRHEUENUCYSPRO 352  
Db 823 CTGCGGCGACTGTGTGAGCGGCGGAGCGGAGCGGGGTGAACATGTTCTGCGCCG 882  
Qy 353 PROARGLEUCYSTRHASPANSGLYILEMETILEATRPASNGLYILEGLUARG 372  
Db 883 GTGCGGCTGTGACGAGACATGGCGGCAATGCGGTGCGCGGAGATGAGCGGTACCG 942  
Qy 373 ALAGLYLEU 375  
Db 943 CGCGGCTG 951  
RESULT 12  
US-09-902-540-503/C  
; Sequence 503, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIORITY FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIORITY FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 503  
; LENGTH: 2582  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-503  
Alignment Scores:  
Pred. No.: 1,62e-50 Length: 2582  
Score: 482.00 Matches: 128  
Percent Similarity: 52.48% Conservative: 52  
Best Local Similarity: 37.32% Mismatches: 131  
Query Match: 22.68% Indels: 32  
Gaps: 9  
DB: 4  
US-10-649-273-2 (1-414) x US-09-902-540-503 (1-2582)  
Qy 38 ILEVALLEUGLYILEGLUTHISERCYASAPSPTRHAIALAVALAVALASPGLUTHR 57  
Db 1525 CTCGCTCCTGAGCTGAAACCTGTGTATGAGACTGCGCGCGCTCGTGAGAGACGCGC 1466

```

QY      58 GlyanValLeuGlyValAlaIleHisSerGlnThrGluValHisLeuValThrGlyGly 77
Db      1465 CGCCGCGGCTGTGTGGATGTGCTTCCACGAGGTGACATCCACGCGGGGTGG 1406
QY      78 IleValProProAlaIleGlnGlnLeuHisValGluValHisLeuValThrGlyGly 97
Db      1405 GTGGTGTCCGAGCTGGCCAGCGCAACCATCTGTCAGTGTGCTGCTCCACGAG 1346
QY      98 AlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleVal 117
Db      1345 GCGCTGACGCGGCGCAACAGAGCTGCGACGAGCTGACCTCATGCTGCGCTCACGTCGCG 1286
QY      118 ProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGln 137
Db      1285 CCGGAGCTCATCGGCGGCTGTGTGGAGTGCAGTGGCCAGAGGCTTGAGCCCTGGGG 1226
QY      138 LeuValValProPheIleProIleHisMetGluAlaHisAlaLeuThrIleArgLeu 157
Db      1225 ACGGGCAAGCCCTTCTGTGGGCGCAACCATCTGAGAGCGCACCTGCTGCGCATCCGGCTG 1166
QY      158 ThrAnlyVal----GluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeu 176
Db      1165 TTGAGAGTGGCGCGGAGCCCGCTTCTGTGGGCTGCTGCTTCCGCGCGGCAACACGAC 1106
QY      177 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLeuSerLeuAspIleAla 196
Db      1105 CTCTACAGAGGTGCAAGCTTACGGGAGTACCGGCTGTGTGGAGCAGCAGCGCAACGCGG 1046
QY      197 ProGlyAspMetLeuAspLeuValAlaArgArgSerLeuIleValHisIleProGlyCys 216
Db      1045 GCCGCGAGGATATGACAAAGACGCTCGCATCTCGGCTG---CCGATACG----- 995
QY      217 SerThrMetSerGlyGlyValAlaIleGluHisLeuAlaValGlnGlyAsn----- 233
Db      994 -----GGTGGCGAGCCCATTCACCATGTTGGCGCAGGAGGAAACCGGAGGCC 947
QY      234 ---ArgPheHisPheAspIleLeuAspProLeuHisHisAlaValAsnCysAspPheSer 252
Db      946 ATCCGCTTC-----CCGCGCGGCTGCGCGGCGCAACTTCCACGCTTC 902
QY      253 PheThrGlyLeuGlnHisValThrAspValIleIleMetValValGluValGluGly 272
Db      901 TTCTCCGCGGTTGAAG-----ACGGCGGTGCTGCACACGCTGCAGAAAGCAGCGC 854
QY      273 IleGluValGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThr 292
Db      853 GTGCGGAGGCGGAGCGCTG-----GCGGATTTGCGCGCTTCCAGGAGGCGC 803
QY      293 MetAlaCysHisLeuValValArgThrHisArgAlaIleLeuPheCysValGlnArgAsp 312
Db      802 GTGGCGGAGCTGCTGTGCAAGAG-----CTGTGGCGCGCGCGCGCGCG 758
QY      313 LeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheVal 332
Db      757 TTG-----GCCACAGACAGTGTGCTGCTGCGCGCGCTGCGCGCACTCGCGG 707
QY      333 IleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysAspPro 352
Db      706 CTGGGCGGACTGTGTGACGCGCGAGCGGAGGCGGGGTTGAACATGTTCTGCCCCG 647
QY      353 ProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArg 372
Db      646 GTGGCGCTGTGCAAGCAATGCGCCCATGATGTGGTGGCGGGTATGAGCGCTATGCGC 587
QY      373 AlaGlyLeu 375
Db      586 CGCGGCGCTG 578

```

RESULT 13  
 US-08-087-797-1  
 ; Sequence 1, Application US/08087797  
 ; Patent No. 554312  
 ; GENERAL INFORMATION:

```

? APPLICANT: Mellors, Alan
? APPLICANT: Lo, Reggie Y.C.
? APPLICANT: Abdullah, Khalid M.
? TITLE OF INVENTION: Pasteurella Haemolytica
? TITLE OF INVENTION: Glycoprotease
? TITLE OF INVENTION: Gene and the Purified Enzyme
? NUMBER OF SEQUENCES: 3
? CORRESPONDENCE ADDRESS:
? ADDRESSER: Bell, Seltzer, Park & Gibson, P.A.
? STREET: 1211 East Morehead Street,
? CITY: Charlotte
? STATE: No. 554312ch Carolina
? COUNTRY: United States
? ZIP: 28214
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/087,797
? FILING DATE: 14-JUL-1993
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Layton, Jr., Samuel G.
? REGISTRATION NUMBER: 22807
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 704 334 2014
? TELEFAX: 704 337 1561
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1315 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 141..1115
? US-08-087-797-1

Alignment Scores:
Pred. No.: 1.19e-50 Length: 1315
Score: 479.00 Matches: 128
Percent Similarity: 48.63% Conservative: 50
Best Local Similarity: 34.97% Mismatches: 157
Query Match: 22.54% Indels: 31
DB: 1 Gaps: 8

US-10-649-273-2 (1-414) x US-08-087-797-1 (1-1315)
QY      9 GlyValPhePheValProSerValArgValValThrGluPheLeuArgSerPheAsnPhe 28
Db      89 GGGGCGCTTCTGCGCCCT-----TTGGTTTCTTAACCTTATTTGACTTC 133
QY      29 HisProGlyThrLeuPheLeuHisValIleValLeuGlyIleGluThrSerCysAspAs 48
Db      134 TCC-----AACTATGCGAATTTTAGGTATTAACCTCTTGATGA 175
QY      48 PThrAlaAlaValAlaValAspGluThrGluValValLeuGlyValAlaIleHisSerGly 68
Db      176 AACCGGTGTGCATTATGATGAGAACAAAGGCTTATGGGCAACAGCTTATAGCCA 235
QY      68 nThrGluValHisLeuValThrGlyGlyIleValProProAlaIleGlnGlnLeuHisAr 88
Db      236 AATTGATATGCAAGCGCGATTACGATGATGATGATGATGATGATGATGATGATGATGAT 295
QY      88 gGluValIleGlnAlaGlyIleValGlnGlnValLeuSerAlaSerGlyValSerProSerAs 108
Db      296 CCGTAAACGTTGCCATTAATTCAGAAAGGCTTAAAGAGGCAATCTGCACACCTCGGA 355

```





Db 552728 GCGGAGCCTTGATATAACAGCAAAATTACTTGACTA---GATTATCA----- 552681  
 QY 218 ThrMetSerGlyGlyValAlaIleGluHisLeuAlaValGlnGlyValAsnArgPheHisIsphe 237  
 Db 552680 -----GGTGGCGCGGACCTTCTCTCTTTCGCGGAAAAGGTACGCCAAATCGCTTTC 552630  
 QY 238 AspIleLysProProLeuHisHisValAlaValAsnCyAspPheSerPheThrGlyLeuGln 257  
 Db 552629 ACATTTCACGTCGCAATGACAGATGCGAGCGCTTGATTTAGTTTCTTGCTTTAAAA 552570  
 QY 258 HisValThrAspLysIleIleMetLysGlyLysGlnGluGly-----IleGluLys 275  
 Db 552569 ACATTTCGCGCAATATGATATCAAGCAATTAATAACAGGCGCAACTGATAGAGCAA 552510  
 QY 276 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCys 295  
 Db 552509 -----ACTTAACCAAGATATGCTTATGCTTTCGCAAGATCGGGGTGGAT 552465  
 QY 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 315  
 Db 552464 ACTCTTGGC-----ATTAATGTAG---CGTCATTGAAAGAA 552429  
 QY 316 GlnAsnAlaValAlaLeuValAlaSerGlyValAlaSerAsnPheTyrIleArgArg 335  
 Db 552428 ACAGGCTATAACGTTTATGATGTCGGGAGGGGTAGCGCAATTAATAACTCGAGAA 552369  
 QY 336 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCyProProProAlaGlu 355  
 Db 552368 ACGCTTGGCGCACTTATATGCAAAATTAGGTGCGCAAGCTTTATCTCAACCTCAATTT 552309  
 QY 356 CysThrAspAsnGlyIleIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGly 374  
 Db 552308 TGTCACGATATATGTCGCGATGATTCCTTACACAGGTTTTCACGTTTAAACAAGGT 552252  
 RESULT 15  
 US-09-643-990A-1/c  
 : Sequence 1, Application US/09643990A  
 : Patent No. 6528289  
 : GENERAL INFORMATION:  
 : APPLICANT: Robert D. Fleischmann  
 : Mark D. Adams  
 : Owen White  
 : Hamilton O. Smith  
 : J. Craig Venter  
 TITLE OF INVENTION: The Nucleotide sequence of  
 the Haemophilus influenzae Rd Genome, fragments  
 thereof, and Uses Thereof  
 NUMBER OF SEQUENCES: 1  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville,  
 STATE: MD  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3 1/2 inch diskette  
 COMPUTER: Dell Pentium  
 OPERATING SYSTEM: MS DOS V6.22  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/643, 990A  
 FILING DATE: 23-Aug-2000  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/487,429  
 FILING DATE: 1995-06-07  
 APPLICATION NUMBER: 08/426,787  
 FILING DATE: 1995-04-21  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kenley K. Hoover  
 REGISTRATION NUMBER: 40,302  
 REFERENCE/DOCKET NUMBER: FBI86PICI

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 301-610-5790  
 TELEFAX: 310-309-8439  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1830121 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-643-990A-1  
 Alignment Scores:  
 Pred. No.: 1,21e-44 Length: 1830121  
 Score: 475.50 Matches: 120  
 Percent Similarity: 50.74% Conservative: 52  
 Best Local Similarity: 35.40% Mismatches: 144  
 Query Match: 22.38% Indels: 23  
 DB: Gaps: 7  
 US-10-649-273-2 (1-414) x US-09-643-990A-1 (1-1830121)  
 QY 39 ValLeuGlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGly 58  
 Db 553208 ATCTTAGGCAATTGAACTTCCTGTGATGAAACGGCGTGCGATTTATGATGAGAAAAA 553149  
 QY 59 AsnValLeuGlyGluAlaIleHisSerGlnThrGluValHisLeuValThrGlyIle 78  
 Db 553148 GGAATTATATGCTAATCACTTATGCTTAACTCAATTGCCGTGATCAGATTATGTTGGTGG 553089  
 QY 79 ValProProAlaIleGlnIleLeuHisArgGluAsnIleGlnArgIleValGlnAla 98  
 Db 553088 GTTCTCGAATTACACATCAGCTATCATATTGCGAAAAACGACCTCTTATTAAGCGCA 553029  
 QY 99 LeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysPro 118  
 Db 553028 TTAGAGAAAGCAATTAAACCGGAGCGAATATGATGATTTCTTATACAGTGGCCT 552969  
 QY 119 GlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeu 138  
 Db 552968 GGGCTTGGTGGCGCATGCTGTCGGGTGCTACGATTCACAGTTCCTTAGCCTTAGCTGG 552909  
 QY 139 LysLysProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeu--- 157  
 Db 552908 AATGTTCTGCGCATATGCTATCATATGAAAGGCGAATTTACTTGCGCCAAATGCTGAT 552849  
 QY 158 ThrAsnLysValGluPheProPheLeuValIleLeuIleSerGlyGlyHisCysLeuLeu 177  
 Db 552848 GACAAATTCACCGCACCTTCTCTTTGCTGCTGTTGATCGGCTGCCACACTCAATTA 552769  
 QY 178 AlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaPro 197  
 Db 552788 GTGCGGTGCGATGATGAGAAAATATGAAGATATGAGAAATCTATGATGATGCTGCT 552729  
 QY 198 GlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSer 217  
 Db 552728 GCGGAAGCCTTGTATTAACAGCAAAATTATGGACTA---GATTATCA----- 552681  
 QY 218 ThrMetSerGlyGlyValAlaIleGluHisLeuAlaValGlnGlyValAsnArgPheHisIsphe 237  
 Db 552680 -----GGTGGCGCGGACCTTCTCTCTTTCGCGGAAAAGGTACGCCAAATCGCTTTC 552630  
 QY 238 AspIleLysProProLeuHisHisValAlaValAsnCyAspPheSerPheThrGlyLeuGln 257  
 Db 552629 ACATTTCACGTCGCAATGACAGATGCGAGCGCTTGATTTAGTTTCTTGCTTTAAAA 552570  
 QY 258 HisValThrAspLysIleIleMetLysGlyLysGlnGluGly-----IleGluLys 275  
 Db 552569 ACATTTCGCGCAATATGATATCAAGCAATTAATAACAGGCGCAACTGATAGAGCAA 552510  
 QY 276 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCys 295  
 Db 552509 -----ACTTAACCAAGATATGCTTATGCTTTCGCAAGATCGGGGTGGAT 552465

```
QY      296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysIleGlnArgAspLeuPro 315
      |||
Db      552464 ACTCTTGCC-----ATTAAATGTAAAG---CGTCATTGAAGA 552429
      |||
QY      316 GlnAsnAsnAlaValLeuValAsnArgIleValAlaSerAsnPheTyrIleArgArg 335
      |||
Db      552428 ACAGGCTATTAAACGTTAGTGTGCGGAGGGGTGAGCGCAATTAATAAACTCCGAGAA 552369
      |||
QY      336 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 355
      |||
Db      552368 ACGCTTGCGCACTTAATGCAAAATTAGGTGCGAAGNGTTTATCTCAACCTCAATT 552309
      |||
QY      356 CysThrAspAsnGlyIleMetIleAlaTyrPheAsnGlyIleGluArgLeuArgAlaGly 374
      |||
Db      552308 TGTACAGATAATGTCGATGATGCTTACACAGGTTTTCGTTTAAACAAGGT 552252
```

Search completed: February 16, 2005, 21:50:50  
Job time : 2493.15 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 18:03:13 ; Search time 828 Seconds

(without alignments)  
2950.230 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1 MLILTKAGVFPKSKRKRVY.....DISKEVGASIKVPLQKMEI 414

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 5378673 seqs, 2950229984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh  
-Q=/cgnt2\_1/USPTO.spool/US10649273/runat\_14022005\_114706\_16497/app\_query.fasta\_1.1429  
-DB=Published Applications NA -OPMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blomsum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR.SCOR=pcr -THR.MAX=100  
-THR.MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10649273@CGN\_1\_1\_1053@runat\_14022005\_114706\_16497  
-NCPU=6 -ICPU=3 -NO MMAP -LARGESUBQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA. \*

1: /cgnt2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgnt2\_6/ptodata/2/pubpna/PTC\_NEW\_PUB.seq.\*  
3: /cgnt2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgnt2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgnt2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgnt2\_6/ptodata/2/pubpna/PTC\_NEW\_PUBCOMB.seq.\*  
7: /cgnt2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgnt2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgnt2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgnt2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgnt2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgnt2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgnt2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
14: /cgnt2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
15: /cgnt2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgnt2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgnt2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgnt2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgnt2\_6/ptodata/2/pubpna/US10F\_NEW\_PUB.seq.\*  
20: /cgnt2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
21: /cgnt2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
22: /cgnt2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	2197	14	US-10-067-443-1
2	2125	100.0	2197	14	US-10-649-273-1
3	2125	100.0	2197	17	US-10-651-722-1
4	2090.5	98.4	1387	14	US-10-067-443-21
5	2090.5	98.4	1387	17	US-10-649-273-21
6	2090.5	98.4	1387	17	US-10-651-722-21
7	2088	98.3	1245	14	US-10-012-140-6
8	2088	98.3	1245	14	US-10-012-140-4
9	1944	91.5	2208	17	US-10-094-720-400
10	1944	91.5	2890	18	US-10-723-986-7447
11	1747	82.2	1416	17	US-10-120-988-177
12	1385	65.2	1526	14	US-10-067-443-23
13	1385	65.2	1526	17	US-10-649-273-23
14	1385	65.2	1526	17	US-10-651-722-23
15	1186.5	55.8	14364	14	US-10-067-443-20
16	1186.5	55.8	14364	17	US-10-649-273-20
17	1186.5	55.8	14364	17	US-10-651-722-20
18	682	32.1	1917	17	US-10-424-599-66417
19	578	27.2	1628	18	US-10-437-963-11249
20	531.5	25.0	4360	14	US-10-081-051-2
21	524	24.7	1146	17	US-10-282-122A-14674
22	502	23.6	1044	17	US-10-282-122A-26972
23	502	23.6	94750	17	US-10-672-787-38
24	494.5	23.3	1032	17	US-10-282-122A-31043
25	492.5	23.2	1026	9	US-09-815-242-7701
26	492.5	23.2	1026	9	US-10-282-122A-30016
27	488.5	23.0	1194	18	US-10-425-115-167432
28	484.5	22.8	1014	9	US-09-815-242-6207
29	484.5	22.8	1014	17	US-10-282-122A-20511
30	484.5	22.8	1014	17	US-10-282-122A-39301
31	483.5	22.8	1014	17	US-10-282-122A-41977
32	482.5	22.7	1020	17	US-10-282-122A-32254
33	481.5	22.7	1014	9	US-09-815-242-9682
34	475.5	22.4	1029	9	US-09-815-242-6946
35	475.5	22.4	1029	17	US-10-282-122A-22020
36	475.5	22.4	1830121	17	US-10-329-670-1
37	475.5	22.4	1830121	18	US-10-158-865-1
38	469	22.1	1023	17	US-10-282-122A-15870
39	468	22.0	371	18	US-10-430-201-3118
40	468	22.0	371	18	US-10-430-201-3119
41	467.5	22.0	1014	17	US-10-282-122A-21633
42	467.5	22.0	1007	17	US-10-282-122A-19220
43	465.5	21.9	1065	17	US-10-282-122A-29584
44	465	21.9	1023	17	US-10-282-122A-31809
45	462.5	21.8	936	17	US-10-282-122A-8315

## ALIGNMENTS

RESULT 1  
US-10-067-443-1  
; Sequence 1, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 2197  
; TYPE: DNA

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-067-443-1

Alignment Scores:
Pred. No.: 2,41e-247 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2 (1-414) x US-10-067-443-1 (1-2197)

QY 1 MetLeuIleLeuThrLySerThrAlaGlyValPhePheLySProSeLySArgLySValTyr 20
Db 231 ATGCTAATCTGACTAGACGACGAGAGGTTTAAACCATCAAAAAGAAAGTTTAT 290
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisGlySILEValLeu 40
Db 291 GAATTTTAAAGAGTTTAAATTTTCATCTCGAAGACATATTTCTTCATATAAATAGATTTG 350
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60
Db 351 GGAATTTGAACTTAGTGTGATGATACAGCAGCTGCTGTGTGTGATGAACTGAAATGTG 410
QY 61 LeuGlyAlaAlaIleHisSerGlnThrGlyValHisGlyLeuThrGlyValPro 80
Db 411 TTGGAGAAAGCAATACATTTCCCAACTGAAAGTTTAAACAGGTGGAGATTGTCT 470
QY 81 ProAlaIleGlnGlnLeuHisArgGlyAsnIleGlnArgIleValGlnGlyAlaLeuSer 100
Db 471 CCACAGCTCAACAGCTTCAACAGAAATATTTCAACAGATGTGTCAAGAAAGCTCTTCT 530
QY 101 AlaSerGlyAlaSerProSerAspLeuSerAlaIleAlaThrThrIleLySProGlyLeu 120
Db 531 GCCAGTGAAGTCTCTCCAGTGAAGCTCTCAGCAATTTGCACTTAAACATTAACAGGACTT 590
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyLeuLeuLyS 140
Db 591 GCTTTAAGCCCTGGAGGTGGGCTTATCATTTAGCTTACAGCTGTGTGAGACAGTTAAAAAG 650
QY 141 ProPheIleProIleHisHisMetGlyAlaHisAlaLeuThrIleArgLeuThrAsnLyS 160
Db 651 CCATTCATTTCCATTCATCATATGAGGCTCATGCACTTACTATTAGGTTGACCAATATA 710
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 711 GTAGAAATTCCTTTTAAATTTTGAATTTCTGAGAGGTCACTGTCTGTGSCATTAGTT 770
QY 181 GlnGlyValSerAspPheLeuLeuGlyLySLeuSerLeuAspIleAlaProGlyAspMet 200
Db 771 CAAGGAGTTTCAGATTTCTGCTTCTTGAAAGTCTTTTGACATAGCACAGGTGACATG 830
QY 201 LeuAspLySValAlaArgArgLeuSerLeuIleLySHisProGlyCysSerThrMetSer 220
Db 831 CTTGACCAAGGTGGCAAGAAAGCTTTCTTTAATAAAACATCCAGATGTCTCCACCATAGCT 890
QY 221 GlyGlyLySAlaIleGlnHisLeuAlaLySglnGlyAsnArgPheHisAspPheIleLyS 240
Db 891 GGTGGGAAAGCCATGAAACATTTGGCCAAACAGAAATAGATTTTCAATTTCACATCAAA 950
QY 241 ProProLeuHisHisAlaIleAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 951 CTTCCCTTGCATCAGCTAAGCTAAGAGGTTTAAACCATCAAAAAGAAAGTTTAT 1010
QY 261 AspLySILEIleMetLySglnGlySglnGlyIleGlyValGlyGlnIleLeuSer 280
Db 1011 GATTAATAATATATGAAAAAGAAAGAAAGGAAAGGATATGAGAAAGGAGCAATCTGTCT 1070
QY 281 SerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLySArg 300
```

```

Db 1071 TCAGCAGCAGACATTCCTCCACAGTACAGCACACATGCAATGTCATCTTGAAAAAG 1130
QY 301 ThrHisArgAlaIleLeuPheCysLySglnArgAspLeuLeuProGlnIleAsnAlaVal 120
Db 1131 ACACATCGGGCTAATCTGTTTGTATAGCAGAGAGCTTGTACCTCAATAATATGACATG 1190
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeu 1340
Db 1191 CTGGTTGCATCTGGTGTGTGTGCGAAGTAACTTATATCCGAGAGCTCTGGAATTTTA 1250
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLySThrAspAsnGly 1360
Db 1251 ACMAAGCAACACAGTGCATTTGTGTGTCTCCACAGACTATGCACTGATTAATGCG 1310
QY 361 IleMetIleAlaITrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 1380
Db 1311 ATTATGATTTGCAATGATGATGTATTTGAAAGACTACGCTGTGGCATTTTACATGAC 1370
QY 381 IleGluGlyIleArgTyrGluProLySProLeuGlyValAspIleSerLySglnVal 1400
Db 1371 ATGAAAGGCAATCGCTATGAAACCAAAATGTCTCTTGAGTACATATCAAAAAGATT 1430
QY 401 GlyGluAlaSerIleLySValProGlnLeuLySMetGluIle 414
Db 1431 GGAAGAGCTTCCATATAAAGTACACAAATTAATAATGAGATA 1472

RESULT 2
US-10-649-273-1
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-649-273-1

Alignment Scores:
Pred. No.: 2,41e-247 Length: 2197
Score: 2125.00 Matches: 414
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2 (1-414) x US-10-649-273-1 (1-2197)

QY 1 MetLeuIleLeuThrLySerThrAlaGlyValPhePheLySProSeLySArgLySValTyr 20
Db 231 ATGCTAATCTGACTAGACGACGAGAGGTTTAAACCATCAAAAAGAAAGTTTAT 290
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisGlySILEValLeu 40
Db 291 GAATTTTAAAGAGTTTAAATTTTCATCTCGAAGACATATTTCTTCATATAAATAGATTTG 350
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60
```

Db 351 GGAATTGAACTAGTGTGATGATACAGACGCTGTGTGGATGAACATGGAATGTG 410  
Qy 61 LeuGIyGluAlaIleHisSerGlnThrGIyValHisLeuLyThrGIyGIyIleValPro 80  
Db 411 TTGGAGAGAAAGCAATACATTCCTCAACCTGAAGTTCAATTAAAAACAGTGGATGTTCCT 470  
Qy 81 ProAlaIaGIInGlnLeuHisArgGIuAenIleGIInArgIleValGIInGIuAlaLeuSer 100  
Db 471 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATGATGACAGAGACTCTTTCT 530  
Qy 101 AlaseerGIyValSerProSerAspLeuSerAlaIleAlaThrThrIleAspProGIyLeu 120  
Db 531 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATGTGCACTAACATAAACAGAGACTT 590  
Qy 121 AlaleuSerLeuGIyValGIyLeuSerPheSerLeuGIInLeuValGIyGIInLeuLySlys 140  
Db 591 GCTTTAAGCTGGAGGTGGCTTATCATTTACCTTACAGCTGGTAGGACAGTTAAAG 650  
Qy 141 ProPheIleProIleHisIleMetGIuAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
Db 651 CCATTCAATCCCATTCATCATATGAGAGGCTCATGCACCTTACTATTAAGTTGACCAATPAA 710  
Qy 161 ValGIuPheProPheLeuValLeuLeuIleSerGIyGIyHisCysLeuLeuAlaLeuVal 180  
Db 711 GTAGAAATTCCTTTTATGTTCTTTGATTTCTGGAGGTCACTGTCTGTGTGGCAATTAGTT 770  
Qy 181 GInGIyValSerAspPheLeuLeuLeuGIyLySerLeuAspIleAlaProGIyAspMet 200  
Db 771 CAAGAGATTCAGATTTCTGCTCTTGAAAGTCTTTGGACATAGCACAGGTGACATG 830  
Qy 201 LeuAspLyValAlaArgArgLeuSerLeuIleLysHisProGIyCysSerThrMetSer 220  
Db 831 CTTCACAAAGGTGGCAGAGAAAGACTTCTTTAATAAACATCCAGAGTCTCCACATGAGT 890  
Qy 221 GIyGIyLysAlaIleGIuHisLeuAlaLysGIInGIyAsnArgPheHisAspAspIleLys 240  
Db 891 GGTGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATTTCATTTTACATCAAA 950  
Qy 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGIyLeuGIInHisValThr 260  
Db 951 CCTCCCTGTGATCATGCTAAATAATGTGATTTTCTTTTACCTGACCTTCAACAGCTTACT 1010  
Qy 261 AspLyIleIleIleMetLyLysGIuLysGIuGIyIleGIuLysGIyGIInIleLeuSer 280  
Db 1011 GATAAATATTAATGAAGAAAGAAAGAAAGAAAGAGATTGAGAAAGGGGCAAAATCCTGCT 1070  
Qy 281 SerAlaIaAspIleAlaAlaThrValGIInHisThrMetAlaCysHisLeuValLysArg 300  
Db 1071 TCAGCAGCAGCATTTGCTGCCACAGTACACACAAATGCAATGCTTGTGAAAGCA 1130  
Qy 301 ThrHisArgAlaIleLeuPheCysLysGIInArgAspLeuLeuProGIInAsnAlaVal 320  
Db 1131 ACACATGGGGCTATTTCTGTTTGTAAAGCAGAGAGCTTGTACTTCAAAATATATGACAGTA 1190  
Qy 321 LeuValAlaSerGIyGIyValAlaSerAsnPheIleIleArgArgAlaLeuGIuIleLeu 340  
Db 1191 CTGTTTCATCTGGTGGTGTGCGCAAGTAATCTTATTCGCGCAGAGCTCTGCAAAATTTTA 1250  
Qy 341 ThrAsnAlaThrGIInCysThrLeuLeuCysProProArgLeuCysThrAsnAsnGIy 360  
Db 1251 ACNAAAGCAACAGTGCATTTGTGTGTCTCTCCACAGCTAATGCACTATATATGTC 1310  
Qy 361 IleMetIleAlaTrpAsnGIyIleGIuArgLeuArgAlaGIyLeuGIyIleLeuHisAsp 380  
Db 1311 ATATGATGATGATGAATGTATTAAGAAAGCTAGCTGTGGGCAATTTTACTATGAC 1370  
Qy 381 IleGIuGIyIleArgIleGIuProLysCysProLeuGIyValAspIleSerLysGIuVal 400  
Db 1371 ATAAAGGCAATCCGCTATGCAACCAAAATGCTCTTGGAGTAGAATATCAAAAGAGTT 1430  
Qy 401 GIyGIuAlaSerIleLysValProGIInLeuLysMetGIuIle 414

Db 1431 GGAGAGCTTCATTAAGTACCAATTAATAAGAGATA 1472  
RESULT 3  
US-10-651-722-1  
; Sequence 1, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; CURRENT FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 2197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (231)..(1472)  
US-10-651-722-1  
Alignment Scores:  
Pred. No.: 2,416-247 Length: 2197  
Score: 2125.00 Matches: 414  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-10-649-273-2 (1-414) x US-10-651-722-1 (1-2197)  
Qy 1 MetLeuIleLeuThrIleThrAlaGIyValPhePheLysProSerLysArgValIleYr 20  
Db 231 ATGCTAATCTTGAATCTGACATGCAAGAGATTTTAAAAACATCAAAAGAAAGTTTAT 290  
Qy 21 GluPheLeuArgSerPheAsnPheHisProGIyThrLeuPheLeuHisLysIleValLeu 40  
Db 291 GAATTTTAAAGAGTTTAAATTTTCACTCTGGAACACTATTTCTTCAATAAATAGATATG 350  
Qy 41 GIyIleGIuThrSerCysAspAspThrAlaAlaValValAspGIuThrGIyAsnVal 60  
Db 351 GGAATTTGAAACTAGTTGTATGATATACAGCAGCTGCTGTGTGGATGAACTGAAATGTG 410  
Qy 61 LeuGIyGluAlaIleHisSerGlnThrGIuValHisLeuLyThrGIyGIyIleValPro 80  
Db 411 TTGGAGAGAAAGCAATACATTCCTCAACCTGAAGTTCAATTAAAAACAGTGGATGTTCCT 470  
Qy 81 ProAlaIaGIInGlnLeuHisArgGIuAenIleGIInArgIleValGIInGIuAlaLeuSer 100  
Db 471 CCAGCAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATGATGACAGAGACTCTTTCT 530  
Qy 101 AlaseerGIyValSerProSerAspLeuSerAlaIleAlaThrThrIleAspProGIyLeu 120  
Db 531 GCCAGTGAAGTCTCTCCAAAGTACCTCTCAGCAATGTGCACTAACATAAACAGAGACTT 590  
Qy 121 AlaleuSerLeuGIyValGIyLeuSerPheSerLeuGIInLeuValGIyGIInLeuLySlys 140  
Db 591 GCTTTAAGCTGGAGGTGGCTTATCATTTACCTTACAGCTGGTAGGACAGTTAAAG 650  
Qy 141 ProPheIleProIleHisIleMetGIuAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
Db 651 CCATTCAATCCCATTCATCATATGAGAGGCTCATGCACCTTACTATTAAGTTGACCAATPAA 710  
Qy 161 ValGIuPheProPheLeuValLeuLeuIleSerGIyGIyHisCysLeuLeuAlaLeuVal 180

Db 711 GTAGATTTCTTTTGTAGTTCTTTGTAGTTCTGAGGTCACGTGCTGTGGCATTAGT 770  
Qy 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyLysMet 200  
Db 771 CAAGAGATTGAGATTTCTCTCTTGAAGGCTTTTGAGATGACACAGTACAG 830  
Qy 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220  
Db 831 CTTGACAAAGGTGGCAAGAGACCTTTCTTAATTAACATCCAGATGCTCCACCATGAT 890  
Qy 221 GlyGlyLysAlaIleGlyHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
Db 891 GGTGGAAAGCCATGAGACATTTGGCCAAACAGAAATAGATTTCATTTGACATCAAA 950  
Qy 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
Db 951 CTTCCCTTGACATGCTTAAATAATGTGATTTCTTTTAACTGAGACTTCACACGTTACT 1010  
Qy 261 AspLysIleIleMetLysLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGly 280  
Db 1011 GATTAATTAATTAATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1070  
Qy 281 SerAlaAlaSerIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300  
Db 1071 TCAGCAGACAGACATTTGCTGCCACAGTACAGACACAAATGGCATGCTTGTGAAGA 1130  
Qy 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
Db 1131 ACACATCGGGCTATCTCTTTTGTGAAGACAGAGACCTGTATCTTCAAAATATGACAGTA 1190  
Qy 321 LeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArgAlaLeuGlnLysLeu 340  
Db 1191 CTGTTGTCATCTGTGTGTGTGCAAGTAACTTCTATTCGCAAGCTTGGAATTTTA 1250  
Qy 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
Db 1251 ACAACGCAACACAGTGCATTTGTGTCTCTCCACAGCTATGACATGATATAGGC 1310  
Qy 361 IleMetIleAlaIleArgGlyIleGlyLysLeuAlaGlyLysGlyLysGlyLysGly 380  
Db 1311 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1370  
Qy 381 IleGlyGlyIleArgGlyGlyLeuProLysCysProLeuGlyValAspIleSerLysGlyVal 400  
Db 1371 ATAGAAGGCATCCGCTATGAACCAAAATGCTCTTGGAGTACATATCAAAAGAGTT 1430  
Qy 401 GlyGlyLysAlaSerIleLysValProGlnLeuLysMetGlnIle 414  
Db 1431 GGAGAGCTTCCATTAAGTACCACAATTAAATAAGAGATA 1472

RESULT 4  
US-10-067-443-21  
; Sequence 21, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 21  
; LENGTH: 1387  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-067-443-21

Alignment Scores:  
Prod. No.: 1,84e-243 Length: 1387  
Score: 2090.50 Matches: 412  
Percent Similarity: 93.85% Conservative: 0  
Best Local Similarity: 93.85% Mismatches: 2  
Query Match: 98.38% Indels: 25  
DB: 14 Gaps: 1

US-10-649-273-2 (1-414) x US-10-067-443-21 (1-1387)

Qy 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
Db 24 ATGCTAATCTTGAACATACTGACAGAGATTTTAAACCAATCAAAAGAAAGTTAT 83  
Qy 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
Db 84 GAATTTTAAAGATTTTAAATTTTCAATCTGAAACCTATTTCTTATATAAAATGATATG 143  
Qy 41 GlyIleGlyThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60  
Db 144 GGAATTAAGAACTAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 203  
Qy 61 LeuGlyGlyAlaIleHisSerGlnThrGlyValHisLeuLysThrGlyIleValPro 80  
Db 204 TTGGGAGAAACATATACATCCCAACTGAAGTTCAATTTAAACAGGTGGATTGCTCT 263  
Qy 81 ProAlaAlaGlnGlnLeuHisArgGlyLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100  
Db 264 CCAGCAGCTCAACAGCTTCAACAGAAATATTCACAGATATACAGAAAGCTCTTTCT 323  
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
Db 324 GCCATGAGATCTCTCCAAAGTAACTCTCAAGATTTGCAACTCAATTAACAGAGACTT 383  
Qy 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
Db 384 GCTTTAAGCTGGAGATGGGCTTATCATTTAGCTTACAGCTGTGAGACAGTTAAAGAA 443  
Qy 141 ProPheIleProIleHisMetGlyAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
Db 444 CCATTCATCCCATTCATATATGAGAGCTCATGACCTTATCTTATGTTGATGACCAAT 503  
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
Db 504 GTGAATTTCTTTTGTAGTTCTTTGATTTCTGAGAGTCACTGCTGTGGCATTTAGT 563  
Qy 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
Db 564 CAAGAGTTTCAATTTTCTGCTTCTTGAAGATCTTTGACATAGACACAGGTGACATG 623  
Qy 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMetSer 220  
Db 624 CTTGACAAAGTGGCAAGAGACCTTTCTTAATTAACATCCAGAGTCTCCACATGAGT 683  
Qy 221 GlyGlyLysAlaIleGlyHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
Db 684 GGTGGAAAGCCATGAGACATTTGGCCAAACAGAAATAGATTTCATTTGACATCAAAA 743  
Qy 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
Db 744 CTTCCCTTGACATGCTTAAATAATGTGATTTCTTTTAACTGAGACTTCACACGTTACT 803  
Qy 261 AspLysIleIleMetLysLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGly 280  
Db 804 GATTAATATATATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 863  
Qy 274 -----GluLys 275  
Db 864 CAGATAAATATTCCTGATTTGCTTAAATAAGTCTCATTTCTGACAGTATGAGAA 923  
Qy 276 GlyGlnIleLeuSerSerAlaAlaSerIleAlaIleAlaThrValGlnHisThrMetAlaCys 295  
Db 924 GGGCAATCTGCTTTCAGCAGCAGACATTTGCTGCCACAGTACAGACACAAATGGCATGT 983

QY 296 HIsleuValIySargThrHisArgAlaIleuPheCylsGlnArgAspLeuPro 315  
 DB 984 CATCTTGGAAGAAAGACATCGGCTATTCGTTTGTGAAGAGAGACATTTGTAACCT 1043  
 QY 316 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPhetYrIleArg 335  
 DB 1044 CAAATAATATGACGTCTGCTGTCATCTGGTGTGTGCAAGTAACCTTCTATATCCGACA 1103  
 QY 336 AlaIeuGluIleleuThrasnAlaThrGlnCysThrIleuLeuCyseProProAspGlu 355  
 DB 1104 GCTCTGGAATTTTACAAACGCAACAGTCACCTTGTGTGCTCTCCCAACACTA 1163  
 QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375  
 DB 1164 TGCACGTAAATAGGATTAATGATTCATGCAATGATTAAGAACTACGTGTGCTTG 1223  
 QY 376 GlyIleleuHisAspIleGluGlyIleArgYrGluProLysCysProleuGlyValAsp 395  
 DB 1224 GGCATTTTACATGACATGAAGGATCCGCTATGAACCAAAATGTCCTTGGAGTAGAC 1283  
 QY 396 IleserIyGluValIleGlyIleAlaSerIleLysValProGlnIleuLysMetGluIle 414  
 DB 1284 ATATCAAAAGAGTTGAGAGGCTTCATTAAGTACCAATTAATAATGAGATG 1340  
 RESULT 5  
 US-10-649-273-21  
 ; Sequence 21, Application US/10649273  
 ; Publication No. US20040043407A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bristol-Myers Squibb Company  
 ; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
 ; FILE REFERENCE: D0073 CNT  
 ; CURRENT APPLICATION NUMBER: US/10/649,273  
 ; CURRENT FILING DATE: 2003-08-27  
 ; PRIOR APPLICATION NUMBER: US 60/266,518  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: US 10/067,443  
 ; PRIOR FILING DATE: 2002-02-05  
 ; PRIOR APPLICATION NUMBER: US 60/282,814  
 ; PRIOR FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 21  
 ; LENGTH: 1387  
 ; TYPE: DNA  
 ; ORGANISM: homo sapiens  
 US-10-649-273-21  
 Alignment Scores:  
 Pred. No.: 1,84e-243 Length: 1387  
 Score: 2090.50 Matches: 412  
 Percent Similarity: 93.85% Conservative: 0  
 Best Local Similarity: 93.85% Mismatches: 2  
 Query Match: 98.38% Indels: 25  
 DB: 17 Gaps: 1  
 US-10-649-273-2 (1-414) x US-10-649-273-21 (1-1387)  
 QY 1 MetIeuIleuThrasnAlaThrGlnCysThrIleuLeuCyseProProAspGlu 20  
 DB 24 ATGCTAATCTTACCTAAGACATGCAAGGTTTTTTTAAACCAATCAAAAGGAAGCTTAT 83  
 QY 21 GluPheLeuArgSerPheAsnPhetIleProGlyThrLeuPheLeuHisIleValIleu 40  
 DB 84 GAATTTTAAAGAACTTTAATTTTCATCTGAAACACATATTTCTCATTAATATGATTTG 143  
 QY 41 GlyIleleuHisAspIleGluGlyIleArgYrGluProLysCysProleuGlyValAsp 395  
 DB 144 GGAATTCAAATCTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 203  
 QY 61 IleserIyGluValIleGlyIleAlaSerIleLysValProGlnIleuLysMetGluIle 414  
 DB 1284 ATATCAAAAGAGTTGAGAGGCTTCATTAAGTACCAATTAATAATGAGATG 1340

DB 204 TTGGAGAAACCAATATATCCCAACTGAAGTTCAATTAATAAAACAGGAGGATTTCT 263  
 QY 81 ProAlaIleGlnIleuHisArgGluAsnIleGlnArgIleValIleGlnIleuValSer 100  
 DB 264 CACAGAGCTAACAGCTTCAACAGAAATATTCAGAAATAGTACAAAGAGCTTTCT 323  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleYrProGlyLeu 120  
 DB 324 GCGAGTGAATCTCTCCAACTGACCTTCCAGCAATTCACCACTAACCAAAACAGACTT 383  
 QY 121 AlaLeuSerIleuGlyValIleuSerPheSerIleuGlnIleuValIleGlnIleuLys 140  
 DB 384 GCTTTAAGCTCGGAGAGGCTTATCATTTAGCTTACAGCTGGTAGACAGTTAAAAAG 443  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaIleuThrIleArgLeuThrAsnLys 160  
 DB 444 CCATTAATCCCATCATCATATGAGAGGCTCATGACCTTACTATAGTTGACCAATTA 503  
 QY 161 ValGluPheProPheLeuValIleuLeuIleSerGlyHisCysLeuLeuAlaLeuVal 180  
 DB 504 GTRGAATTTCTTTTATGATTTCTTTGATTTCTGGAGTCACTGTCTTTGGCATTTAGTT 563  
 QY 181 GlnGlyValSerAspPheLeuLeuGlyLysSerIleuAspIleAlaProGlyAspMet 200  
 DB 564 CAAGAGTTTCAGATTTTCTGCTTCTTGAAAGTCTTTGACATAGCACAGGAGATG 623  
 QY 201 LeuAspLysValAlaArgArgLeuSerIleuIleLysHisProGluCysSerThrMetSer 220  
 DB 624 CTTGACAGGTGCAAGAAAGACTTTCTTAATTAACATCCAGAGTCTCCACACATGAGT 683  
 QY 221 GlyIleLysAlaIleGlnIleuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 DB 684 GGTGGAAAGCCATAGACATTTGGCCAAACAGAAATGATTTTATTTGACATCA 743  
 QY 241 ProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGlyLeuGlnIleValThr 260  
 DB 744 CCGCTTCATCATGACTAAATTTGATTTTCTTTTACGTGAGCTTCAACAGCTTACT 803  
 QY 261 AspLysIleIleMetLysLysGluLysGluGlyIle----- 273  
 DB 804 GATTAATAATATATGAAAAAGAAAAAGAAAGATATTTCTTAATAGTAAGTTGAA 863  
 QY 274 -----GluLys 275  
 DB 864 CAGATAATATTCCTGATTTGCTTAAATAAGCTGCTCATTTCTGACGATAGAGAG 923  
 QY 276 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 295  
 DB 924 GGGCAATCTCTGCTTCAAGCAGACATTTGCTGCCAGCTACGACACATGCGCATGT 983  
 QY 296 HIsleuValIySargThrHisArgAlaIleuPheCylsGlnArgAspLeuPro 315  
 DB 984 CATCTTGGAAGAAAGACATCGGCTATTCGTTTGTGAAGAGAGACATTTGTAACCT 1043  
 QY 316 GlnAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPhetYrIleArg 335  
 DB 1044 CAAATAATGACGTCTGCTGTCATCTGGTGTGTGCAAGTAACCTTCTATATCCGACA 1103  
 QY 336 AlaIeuGluIleleuThrasnAlaThrGlnCysThrIleuLeuCyseProProAspGlu 355  
 DB 1104 GCTCTGGAATTTTACAAACGCAACAGTCACCTTGTGTGCTCTCCCAACACTA 1163  
 QY 356 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 375  
 DB 1164 TGCACGTAAATAGGATTAATGATTCATGCAATGATTAAGAACTACGTGTGCTTG 1223  
 QY 376 GlyIleleuHisAspIleGluGlyIleArgYrGluProLysCysProleuGlyValAsp 395  
 DB 1224 GGCATTTTACATGACATGAAGGATCCGCTATGAACCAAAATGTCCTTGGAGTAGAC 1283  
 QY 396 IleserIyGluValIleGlyIleAlaSerIleLysValProGlnIleuLysMetGluIle 414  
 DB 1284 ATATCAAAAGAGTTGAGAGGCTTCATTAAGTACCAATTAATAATGAGATG 1340



```
RESULT 6
US-10-651-722-21
; Sequence 21, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-651-722-21

Alignment Scores:
Pred. No.: 1,84e-243 Length: 1387
Score: 2090.50 Matches: 412
Percent Similarity: 93.85% Conservative: 0
Best Local Similarity: 93.85% Mismatches: 2
Query Match: 98.38% Indels: 25
DB: 17 Gaps: 1

US-10-649-273-2 (1-414) x US-10-651-722-21 (1-1387)
QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20
Db 24 ATGCAATCTTGCATCAAGAGCTCAGAGAGTTTATTAACATCAAAAGAAAGTTTAT 83
QY 21 GluPheLysArgSerPheAsnDheHisProGlyThrLysPheLysHisLysIleValLeu 40
Db 84 GAATTTTAAAGAAAGTTTATTAATTTTCACTGAAACACATATTTCTTCAATAAATAGATTTG 143
QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValValAspGluThrGlyValAsnVal 60
Db 144 GGAATTGAAACTGATGTGATGATACAGACGCTGCTGTGATGAAACTGGAATATGTG 203
QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLysLysThrGlyIleValPro 80
Db 204 TTGGGAGAGCAATACATTCCTCAAACTGAAGTTCAATTTAAAAACAGGTGGATTTCTCT 263
QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnIleLysSer 100
Db 264 CCAGAGCTCAACACACTTCACAGAGAAATATTTCAACGAATAGTACAAAGAGCTTTTCT 323
QY 101 AlaSerGlyValSerProSerAspLysSerAlaIleAlaThrThrIleLysProGlyLeu 120
Db 324 GCCAGTGAAGCTCTCCCAAGTACCTTCAGCAATTTGCAACTAACATTAACCAAGACTT 383
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140
Db 384 GCTTTAAGCTGGAGAGTGGCTTATTCATTTAGCTTACAGCTGTGATGAGACGTTAAAAAG 443
QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160
Db 444 CCATTCATTCCTCCATTCATATGAGGCTCAATGACCTTACTATTAGTTGACCAATATA 503
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaVal 180
Db 504 GTAGAAATTCCTTTTATGTTCTTTGATTTCTGAGAGTCACTGCTGTTGGCATTAAGTT 563
QY 181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200
```

```
Db 564 CAAGAGTTTCAGATTTTCCTGCTTCTTGGAAGCTCTTGACATAGACCAAGGTGACATG 623
QY 201 LeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220
Db 624 CTTGACAGGTGCAAGAAAGACTTCTTTATATTAACATTCACAGAGCTTCACCATAGT 683
QY 221 GlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
Db 684 GGTGGGAAAGCCATAGACATTTGGCCAAACAAAGAAATGATTTCAATTTGACATCAAA 743
QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 744 CCGCCCTGCATCATCTATAAATGTGATTTCTTTACTGACACTTCAACAGCTTACT 803
QY 261 AspLysIleIleMetLysLysGlyLysGluGlnGlyIle----- 273
Db 804 GATAAATAATATATGAAAAAGAAAAAGAAAGATATATTTCTTAATAGTAAAGTTGAA 863
QY 274 -----GluLys 275
Db 864 CAGATAAATATTCCTGAGATTTGCTTAAATAAGCTGCTCATTTCTGCAGGTATGAGAG 923
QY 276 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 295
Db 924 GGGCAATCCTGCTTACAGACAGACATTTGCTGCACAGTACAGCACAAATGGCATGT 983
QY 296 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 315
Db 984 CATCTTGAAAAAGAACACATCGGCTATTTCTTTGTGTAAGACAGACAGCTGTATCCT 1043
QY 316 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaLysAsnPheThrIleArgArg 335
Db 1044 CAAATAATGCACTACTGCTGTGATCTGTGTGCTGCTGCAAGTAACTTCTATTCGCGAGA 1103
QY 336 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 355
Db 1104 GCTCTGAAATTTTAAACAAACGCAACAGTGCATTTGTTGTGTCCTCCACAGACTA 1163
QY 356 CysThrAspAsnGlyIleMetIleAlaIleArgGlyIleGluArgLeuAlaGlyLeu 375
Db 1164 TGCACGTATATGCACTTATGATTTGATGCAATGGAATGGAATTAAGACATACGTGTGCTTG 1223
QY 376 GlyIleLeuHisAspIleGluGlyIleArgGlyLeuProLysCysProLeuGlyValAsp 395
Db 1224 GGCATTTTACATACATTAAGAGCAATCCGCTTAAGAACAAATGTCTCTTGGAGTGAC 1283
QY 396 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
Db 1284 ATATCAAAAGAAAGTTGAGAGAGCTTCATTAAGTACCAATTTAAAAATGGAGATA 1340

RESULT 7
US-10-012-140-6
; Sequence 6, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Liebermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
```

LENGTH: 1245  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-10-012-140-6

## Alignment Scores:

Pred. No.: 3,13e-243 Length: 1245  
 Score: 2088.00 Matches: 407  
 Percent Similarity: 99.03% Conservative: 3  
 Best Local Similarity: 98.31% Mismatches: 4  
 Query Match: 98.26% Indels: 0  
 DB: 14 Gaps: 0

US-10-649-273-2 (1-414) x US-10-012-140-6 (1-1245)

QY 1 MetLeuIleuThrLysThrAlaGlyValPhePheLysProSerLysValTyr 20  
 DB 1 ATGCTAATCTTGACTTAAGACGTGACAGAGCTTTTAAACATCAAAAAGAAAGTTTAT 60  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
 DB 61 GAATTTTAAAGAGTTTAAATTTTCTGGAACACTATTTCTTCATATAAATGATTG 120  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60  
 DB 121 GGAATTGAACTAGTGTGATGATACACAGCTGCTGTGTGATGAACTGGAATGTG 180  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLeuLysThrGlyGlyIleValPro 80  
 DB 181 TTGGAGAAAGCAATACATTCCTCAACCTGAAGTCTTTTAAACAGGTGGATGTTCT 240  
 QY 81 ProAlaIleGlnIleuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 241 CCACAGCTCAACAGCTTCAAGAAAATATTCACGAATAGTACAAAGACCTCTTCT 300  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
 DB 301 GCCAGTGAAGCTCTCCCAAGTACCTCTCAGCAATTGCAACATCAATAAACAGACTT 360  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLysLys 140  
 DB 361 GCTTAAAGCCCTGGAGTGGCTTATCATTTAGCTTACAGCTGTGTGACAGATTAAAG 420  
 QY 141 ProPheIleProIleHisHisMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
 DB 421 CCATTCAATCCCATTCATCATATGAGGCTCATGACTTACTATTAGGTGACAAATAA 480  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 DB 481 GTAGAAATTTCTTTTATGATTCTTTGATTTCTGAGGCTCACTGCTGTGGCATTTAGT 540  
 QY 181 GlnGlyValSerAspPheLeuLeuGlnLysSerLeuAspIleAlaProGlyAspMet 200  
 DB 541 CAAGGAGATTTCAGATTTTCTGCTTCTTGAAAGCTTTTGACATATGACACAGTACATG 600  
 QY 201 LeuAspLysValAlaAlaGArgLeuSerLeuIleLysHisProGluCysSerThrMetSer 220  
 DB 601 CTGACCAAGGTGGCAGAGAGACTTTCTTAAATAAACATCCAGATGCTCCCACTAGAGT 660  
 QY 221 GlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 DB 661 GGTGGGAAAGCCATAGAACATTTGGCCAAACAGAGAAATAGATTTCATTTGACATCAAA 720  
 QY 241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 DB 721 CTTCCCTTGATCATCTGTAATAATTTGATTTTCTTTTACTGACCTTCAACACCTTACT 780  
 QY 261 AspLysIleIleMetLysLeuGluLysGluGluGlyIleGlyLysGlyGlnIleLeuSer 280  
 DB 781 GATTAATAATATATGATTAACGCTAACTAGAGAGAGATTTGAAAGAGGCAATCTGTCT 840  
 QY 281 SerAlaIleAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArg 300

DB 841 TCAGCAGACAGACATTCCTGCCACAGTACAGACACACATGCGATCTTGTGAAAAA 900  
 QY 301 ThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320  
 DB 901 ACACATCGGGCTATCTGTTTGTATGACAGAGAGACTGTACTCAAAATTAATGACGTA 960  
 QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeu 340  
 DB 961 CTGGTTCATCTGCTGTGTGTGCGCAATTACTTATATCGCAGAGCTCTGAAATTTTA 1020  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360  
 DB 1021 ACAACGCAACACAGTCACTTGTGTGTCTCTCCAGACTATGCACTGATTAAGGC 1080  
 QY 361 IleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1081 ATTATATTTGCAATGAGATGGTATTTAAAGACTAGCTGGCTTGGCATTTTAAACAGAC 1140  
 QY 381 IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
 DB 1141 ATGGAAGCATCGCTATGACCAAAATGCTCTTGGAGTACATATCAAAAGAGTT 1200  
 QY 401 GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 DB 1201 GGAGAGCTTCCATATAAGTACACACATTAATAATGAGATA 1242

RESULT 8  
 US-10-012-140-4  
 ; Sequence 4, Application US/10012140  
 ; Publication No. US20030009017A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Leibny, Kevin R.  
 ; APPLICANT: Kapeller-Liebermann, Rosana  
 ; APPLICANT: Glucksmann, Marcia A.  
 ; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
 ; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
 ; FILE REFERENCE: 381552004900  
 ; CURRENT APPLICATION NUMBER: US/10/012,140  
 ; PRIOR FILING DATE: 2001-11-08  
 ; PRIOR APPLICATION NUMBER: 60/246,768  
 ; PRIOR FILING DATE: 2000-11-08  
 ; PRIOR APPLICATION NUMBER: 60/246,772  
 ; PRIOR FILING DATE: 2000-11-08  
 ; PRIOR APPLICATION NUMBER: 60/249,185  
 ; NUMBER OF SEQ ID NOS: 49  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1820  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (146)...(1390)  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1820)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; US-10-012-140-4

Alignment Scores:  
 Pred. No.: 5.69e-243 Length: 1820  
 Score: 2088.00 Matches: 407  
 Percent Similarity: 99.03% Conservative: 3  
 Best Local Similarity: 98.31% Mismatches: 4  
 Query Match: 98.26% Indels: 0  
 DB: 14 Gaps: 0

US-10-649-273-2 (1-414) x US-10-012-140-4 (1-1820)

QY 1 MetLeuIleuThrLysThrAlaGlyValPhePheLysProSerLysValTyr 20  
 |||

```

Db 146 ATGCTAATCTTGACCTAGACGACGAGATTGTTTAAACATCAAAAAGAAAGTTTAT 205
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
Db 206 GAATTTTAAAGAGTTTAAATTTTCACTCGAACAACATATTTCTTCATATAAATAGATTG 265
QY 41 GlyIleGluThrSerCysAspAspThrIleAlaValAspGluThrGlyAsnVal 60
Db 266 GGAATTTGAAACTAGTTGATGATGATACAGACGCTGCTGCTGCTGATGAAATCGAAATG 325
QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisIleuLeuThrGlyGlyIleValPro 80
Db 326 TTGGAGAGAGCAATTCATTTCCCAACTGAAAGTTTCAATTTTAAAAACGCTGGATGTTTCT 385
QY 81 ProAlaIleGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGlnIleuLeuSer 100
Db 386 CCAAGAGCTCAACAGCTTCAACAGAAATATTCAACGAATAGTAAAGAGCTCTTCT 445
QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleIleProGlyLeu 120
Db 446 GCGAGTGAAGTCTCTCCAAAGTACCTCTGACGAATTCGACATACCAATAAACGAGACTT 505
QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuIleVal 140
Db 506 GCTTTAAGCTGGGAGTGGGCTTATCATTTAGCTTACAGCTGGTAGACAGTTAAAAAAG 565
QY 141 ProPheIleProIleHisIleMetGlnAlaHisAlaLeuThrIleArgLeuThrAsnVal 160
Db 566 CCATTCAATCCATTCATCATATGAGAGGCTCATGACCTTACATTAAGTTGATGACCAATTA 625
QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180
Db 626 GTAGAAATTTCTTTTATGATCTTTGATTTCTGAGAGTCACTGCTGTTGGCATTAAGTT 685
QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIleValSerLeuAspIleAlaProGlyAspMet 200
Db 686 CAAGGAGTTTCAGATTTTCTGCTCTTGGAAGAGTTTGATGACATAGCAACAGGTGACATG 745
QY 201 LeuAspIleValAlaArgLeuSerLeuIleValHisProGluCysSerThrMetSer 220
Db 746 CTTCAGCAAGGTGGCAAGAGACTTCTTTAAATTAACATCCAGAGCTCCACCAAGAGT 805
QY 221 GlyGlyIleValAlaIleGlnHisLeuAlaIleValGlnGlyAsnArgPheHisIleVal 240
Db 806 GGTGGGAAGGCAATGAACTATTTGGCCAAACAGAAATAGATTTTCAATTTGACATCAAA 865
QY 241 ProProLeuHisIleAlaIleValAsnGlyAspPheSerPheThrGlyLeuGlnHisValThr 260
Db 866 CTTCCCTTGACATCATGCTAAATAATGATTTTCTTTTACAGACTTCAACACGTTACT 925
QY 261 AspIleValIleMetIleValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 280
Db 926 GATAAAAATTAAGAAACAGGAAACAGGAAAGGATTAAGAAAGGGAATCCCTGCT 985
QY 281 SerAlaIleAspIleAlaIleAlaThrValGlnHisIleThrMetAlaCysHisIleValIleVal 300
Db 986 TCAGACAGACATCTGCTCCACAGTACGACACATGACATGCTGCTGGAAGAAAG 1045
QY 301 ThrHisArgAlaIleLeuPheCysIleGlnArgAspLeuProGlnAsnAsnAlaVal 320
Db 1046 ACAATCGGGCATATCTGTTTGTAAAGCAGAGACTTGTACCTCAAAATATATGACATG 1105
QY 321 LeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArgAlaLeuGluIleLeu 340
Db 1106 CTGGTTCATCTGGTGTCTCCAGAGTAACTTCATATCCGACAGAGCTCGGAAATTTTA 1165
QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGly 360
Db 1166 ACAAAAGCAACAGTGCATTTGTGTCTCTCCCAAGACTATGACACTGATTAATGAGC 1225
QY 361 IleMetIleAlaArgAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
Db 1226 ATTATGATTCATGATGATGATGATTAAGACATACGCTGCTGGCATTTTACATGAC 1285

```

```

QY 381 IleGluGlyIleArgGlyIleProGlyCysProLeuGlyValAspIleSerIleGlyVal 400
Db 1286 ATGAAAGCATCGCATATGAAACCAAAATCTCTTGGAGTAGACATATCAAAAGAGTT 1345
QY 401 GlyIleValSerIleValValProGlnLeuIleValMetGluIle 414
Db 1346 GGAAGAGCTTCATTAAGATACCAACAATTAAGATTAAGATTAAGATTAAGATTA 1387

RESULT 9
US-10-094-749-400
; Sequence 400, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YUKI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OR INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094, 749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350, 435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 400
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-094-749-400

Alignment Scores:
Pred. No.: 2 578-225 Length: 2208
Score: 1944.00 Matches: 386
Percent Similarity: 93.24% Conservative: 0
Beet Local Similarity: 93.24% Mismatches: 4
Query Match: 91.48% Indels: 24
DB: 17 Gaps: 1

US-10-649-273-2 (1-414) x US-10-094-749-400 (1-2208)
QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLeuProSerIleArgIleValThr 20
Db 344 ATGCTAATCTTGACCTAGACGACGAGATTGTTTAAACATCAAAAAGAAAGTTTAT 403
QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40
Db 404 GAATTTTAAAGAGTTTAAATTTTCACTCGAACAACATATTTCTTCATATAAATAGATTG 463
QY 41 GlyIleGluThrSerCysAspAspThrIleAlaValAspGluThrGlyAsnVal 60
Db 464 GGAATTTGAAACTAGTTGATGATGATACAGACGCTGCTGCTGCTGATGAAATCG 523
QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisIleuLeuThrGlyGlyIleValPro 80
Db 524 TTGGAGAGAGCAATTCATTTCCCAACTGAAAGTTCAATTTAAAAACAGGTGGATTTGCTCT 583

```

QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 584 CCAGAGCTCAACAGCTTCAGAGAAATATTCACAGAAATGTCACAAAGCTCTTCT 643  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120  
 DB 644 GCCAGTGAAGTCTCTCAAGTGAACCTTCAGCAATGCAATCACTAAACAGAGACTT 703  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLeuVal 140  
 DB 704 GCTTAAAGCTGGAGGAGGCTTACATTTACCTTACAGCTGTAAGAGAGATTAAAAAG 763  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeu 160  
 DB 764 CCATTCAATCCATTCATCATATGAGGCTCATGCACTTACTATTAGTTGACCAATTA 823  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 DB 824 GTAGAAATTCCTTTTATTTAGTTCTTTGATTTCTGGAGGTCACGTCTGTGGCATTAGTT 883  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMet 200  
 DB 884 CAAGAGATTCAAGATTTCTGCTTCTTGAAAGCTTTTGACATAGCACAGGTGACATG 943  
 QY 201 LeuAspLeuValAlaArgArgLeuSerLeuIleLeuHisArgGluCysSerThrMetSer 220  
 DB 944 CTTTACAAAGGTGGCAAGAAAGCTTCTTTAATTAACATCCAGAGGCTCCACAGTACT 1003  
 QY 221 GlyGlyValAlaIleGlnHisLeuAlaLeuGlnGlyAsnArgPheHisAspAspIleLeu 240  
 DB 1004 GGTGGAAAGCATATGAACATTTGGCCAAACAAAGAAATAGATTTCATTGACATCAAA 1063  
 QY 241 ProProLeuHisHisAlaLeuAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260  
 DB 1064 CCTCCCTTGACATGCTTAAATATGTAATTTCTTTTAACTGACCTTCAACACCTTACT 1123  
 QY 261 AspLeuValIleIleMetLeuValGlnLeuGlnGlyGlyIleGlnLeuGlyGlnIleLeuSer 280  
 DB 1124 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1183  
 QY 281 SerAlaAlaSerIleAlaAlaThrValGlnHisThrMetAlaCysHisIleValIleValArg 300  
 DB 1184 TCACAGAGAGCATTTGCTGCCAGTACAGCAACAATGGCATGTCATCTTGTGAAGAA 1243  
 QY 301 ThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320  
 DB 1244 ACAATCGGGCTATTTCTTTTGTATGAGAGAGACTTGTATCTCAAAATTAATGCAATA 1303  
 QY 321 LeuValAlaSerGlyValAlaAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeu 340  
 DB 1304 CTGGTTCATCTGGTGGTGGTGGCAAGTATCTGTATCCGCAAGCTCTGGAATTTTAA 1363  
 QY 341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgArgLeuCysThrAspAsnGly 360  
 DB 1364 ACAAAACCAACAGAGTCACTTTGTTGTCTCTCCCAAGCTATGCACTGATATATGCC 1423  
 QY 361 IleMetIleAlaTPAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380  
 DB 1424 ATATATGATTCGA-----TGATGTCCTCTTGAAGTATGACATATCAAAAGAGTT 1435  
 QY 381 IleGlnGlyIleArgGlyArgGluProLeuCysProLeuGlyValAspIleSerLeuGluVal 400  
 DB 1436 -----TGATGTCCTCTTGAAGTATGACATATCAAAAGAGTT 1471  
 QY 401 GlyGluAlaSerIleLeuValProGlnLeuLeuMetGluIle 414  
 DB 1472 GGAGAGCTTCATTAAGTACCAACAATTAATAAATGAGATA 1513  
 RESULT 10  
 ; US-10-723-860-7447  
 ; Sequence 7447, Application US/10723860  
 ; Publication No. US20040253606A1

; GENERAL INFORMATION:  
 ; APPLICANT: Aziz, Natasha  
 ; APPLICANT: Ginsburg, Wendy M.  
 ; APPLICANT: Zlotnick, Albert  
 ; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
 ; FILE REFERENCE: 05882.0193.NPUS01  
 ; CURRENT APPLICATION NUMBER: US/10/723,860  
 ; PRIOR FILING DATE: 2003-11-26  
 ; PRIOR APPLICATION NUMBER: 60/429,739  
 ; PRIOR FILING DATE: 2002-11-26  
 ; NUMBER OF SEQ ID NOS: 8393  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 7447  
 ; LENGTH: 2890  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (646)..(657)  
 ; OTHER INFORMATION: n is a, c, g, or t  
 ; US-10-723-860-7447  
 Alignment Scores:  
 Pred. No.: 3,93e-225 Length: 2890  
 Score: 1944.00 Matches: 386  
 Percent Similarity: 93.24% Conservative: 0  
 Best Local Similarity: 93.24% Mismatches: 4  
 Query Match: 91.48% Indels: 24  
 DB: 18 Gaps: 1  
 US-10-649-273-2 (1-414) x US-10-723-860-7447 (1-2890)  
 QY 1 MetLeuIleLeuThrIleThrAlaGlyValPhePheLeuSerIleValArgValIle 20  
 DB 1001 ATGCTAATCTTGAAGTAAAGTGAAGGATTTTAAACATCAAAAGAAAGTTAT 1060  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValLeu 40  
 DB 1061 GAATTTTAAAGATTTTAAATTTTCAATCTCGAAGCACTATTTCTCAATAAATATGATAATG 1120  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValAlaAspGluThrGlyAsnVal 60  
 DB 1121 GGAATTTGAACATGATTTGATATGATACAGAGCTGCTGTGTGATGAATGGAATG 1180  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisIleLeuThrGlyGlyIleValPro 80  
 DB 1181 TTGGAGAGAGCAATACATTTCCAAACTGAAGTTCATTAAACAGGTGGATTTCTCT 1240  
 QY 81 ProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSer 100  
 DB 1241 CCAGAGCTCAACAGCTTCACAGAGAAATATTCACAGAAATATGACAAAGCTCTTCT 1300  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLeuProGlyLeu 120  
 DB 1301 GCCAGTGAAGTCTTCCAAAGTACCTTCAGCAATGCAATCACTAAACAGAGACTT 1360  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLeuVal 140  
 DB 1361 GCTTAAAGCTGGAGTGGGCTTATCATTTAGCTTCAAGCTGTATGACAGCTTAAAAAG 1420  
 QY 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeu 160  
 DB 1421 CCATTCAATCCATTCATCATATGAGGCTCATGCACTTACTATTAGTTGATCAATTA 1480  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 DB 1481 GTAGAAATTCCTTTTATTTAGTTCTTTGATTTCTGGAGGTCACGTCTGTGGCATTAGTT 1540  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMet 200  
 DB 1541 CAAGAGTTTCAGATTTTCTGCTTCTTGAAAGTCTTTTGACATATGACCAAGGTGACATG 1600

```

QY      201  LeuAspLySValAlaArgArgLeuSerLeuIleuLysHisProGluCySerThrMetSer 220
DB      1601 CTGACAAAGTGGGAGAAAGACTTCCTTAAATAAATCCAGAGTGCCTCCACCATGACT 1660
QY      221  GlyGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240
DB      1661 GGTGGAAAGCCATGAAACATTTGGCCAAACAAAGAAATAGATTTCATTTGACATCAAA 1720
QY      241  ProProLeuHisHisAlaLysAsnCySerAspPheSerPheThrGlyLeuGlnHisValThr 260
DB      1721 CTTCCCTTGACATCAAGCTTAAATAATTTGATTTTCTTTTACCTGACCTTCAACACGTTTACT 1780
QY      261  AspLysIleIleMetLysLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLys 280
DB      1781 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1840
QY      281  SerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
DB      1841 TCACACAGACAGATGCTGCCACAGTACAGCACACAAATGGCATGTCATCTTGAAAAAG 1900
QY      301  ThrHisArgAlaAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal 320
DB      1901 ACACATCGGCTATTCGTGTTTGTAGACAGAGACCTGTTACCTCAAAATTAATGCAGTA 1960
QY      321  LeuValAlaSerGlyGlyValAlaAsnAsnPheTyrIleArgArgAlaLeuGlnIleLeu 340
DB      1961 CTGTTGTCATCTGGTGGTGGTGGCAGTAATCTTGTATCCGACAGCTCTGAAAAATTTTA 2020
QY      341  ThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly 360
DB      2021 ACAACGCAACACAGTGCATTTGTTGTCTCTCCACAGCTATGACATGATATGAGC 2080
QY      361  IleMetIleAlaTyrPheAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAsp 380
DB      2081 ATTATGATGTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2092
QY      381  IleGluGlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGlyVal 400
DB      2093 -----TGATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2128
QY      401  GlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414
DB      2129 GGAGAGCTTCATTAAGTACCAATTTAAAAAGGAGATA 2170

RESULT 11
US-10-120-988-177
; Sequence 177, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Kyle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyan
; APPLICANT: Wang, Dunrui
; APPLICANT: Dimanac, Radoje T.
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/10/120,988
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pc_fl_genes Version 2.0
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205) .. (1305)
US-10-120-988-177

```

## Alignment Scores:

```

Pred. No.: 1,199-201 Length: 1416
Score: 1747.00 Matches: 340
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.71% Mismatches: 0
Query Match: 82.21% Indels: 0
DB: 17 Gaps: 0

```

US-10-649-273-2 (1-414) x US-10-120-988-177 (1-1416)

```

QY      74  LysThrGlyGlyIleValProProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArg 93
DB      280 AGAACAAGGGGAGATGTTCTCCAGACGCTCAACAGCTTCACAGAAAAATATATCAACGA 339
QY      94  IleValGlnGlnAlaLeuSerAlaSerGlyValLysProSerAspLeuSerAlaIleAla 113
DB      340 ATATGACAAAGAGCTTTTCTGCAAGTGAAGTCTCCCAAGTACCTTCAGCAAAATGCA 399
QY      114  ThrThrIleLysProGlyLeuAlaLeuSerLysGlyValGlyLysSerPheSerLeuGln 133
DB      400 ACTACCATTAATAACAGAGACTTCTTAACTGGAGTGGGCTTATCATTTACTTACG 459
QY      134  LeuValGlyGlnLeuLysLysPheProPheIleProIleHisHisMetGluAlaHisAlaLeu 153
DB      460 CTGTATGACAGATTAATAAAGCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 519
QY      154  ThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIleSerGlyGly 173
DB      520 ACTATTAAGGTATCAACAAATTAAGTAATTTCTTTTAAAGTTCTTTGATTTTGAGGT 579
QY      174  HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu 193
DB      580 CACTGTCCTTGGCATTTACTTCAAGAGATTTCAGATTTCTGCTTCTTGGAAAGCTTTTG 639
QY      194  AspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHis 213
DB      640 GACATATGACCAAGGTGACATGCTTGCAAGGTGGCAAGAAGACTTCTTAAATAAATCAT 699
QY      214  ProGluCysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsn 233
DB      700 CCAGAGTCTCCACCATGAGTGGTGGAAAGCATTTGCAATTTGGCAAAACAGAAAT 759
QY      234  ArgPheHisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPhe 253
DB      760 AGATTTCATTTGACATCAACCTCCCTTGACATGCTTAATAATTTGATTTTCTTTT 819
QY      254  ThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIle 273
DB      820 ACTGAGCTTCAACAGCTTACTGATTAATAATTAATTAATAAAGGAAAAAGGAAAGTAT 879
QY      274  GlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMet 293
DB      880 GGAAGGGGCAATCTCTGCTTCAGCAGCAGCATTTGCTCCACAGTACGACACATG 939
QY      294  AlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 313
DB      940 GATGTCATCTTGTAAGAAAGAACACATCGGCTATTTCTGTTTGTAGACAGAGACTTG 999
QY      314  LeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaAlaSerAsnPheTyrIle 333
DB      1000 TTACCTCAAAATTAATGCACTACTGCTGATGCTGGTGGTGGCAAGTAACTTCTATATC 1059
QY      334  ArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPro 353
DB      1060 CCAGAGCTCTGGAATTTTAAACAACGACACATGCGCTTTGTTGTCCTCTCTCTCC 1119
QY      354  ArgLeuCysThrAspAsnGlyIleMetIleAlaTyrAsnGlyIleGlnArgLeuArgAla 373
DB      1120 AGACTATGACATGATATATGCAATTAATGATGATGCAAGTAAATGTAACATACGCT 1179
QY      374  GlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLysCysProLeuGly 393
DB      1180 GGTGGGCAATTTTACATGACATAGAGGCAATCCGCTATGAAACCAAAATCTCTTGA 1239

```

QY 394 ValaSerIleSerIleGluValGluValaSerIleLeuValProGluLeuIleuIleuMetGlu 413  
DB 1240 GTAGACATATCAAAAGAAAGTTGGAGAGAGCTTCATTAAGTACCAATTAAGAGAG 1299

QY 414 Ile 414  
DB 1300 ATA 1302

RESULT 12  
US-10-667-443-23  
; Sequence 23, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-667-443-23

Alignment Scores:  
Pred. No.: 1,49e-157 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 65.18% Indels: 0  
DB: 14 Gaps: 0

US-10-649-273-2 (1-414) x US-10-667-443-23 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeuValGluPheProPheLeuVal 167  
DB 1 ATGAGAGCTCATGCACTTAATAATGAGTTGACCAATTAAGTAAATTTCTTTTAGTT 60

QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187  
DB 61 CTTTGAATTTCTGGAGGTCACTGTCTGTGGCATTAAGTTCAAGAGTTTCAGATTTTCTG 120

QY 188 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArg 207  
DB 121 CTTCTTGAAAGCTTTTGACATGACACAGGTGACATGCTTGACAAAGTGGCAAGAAGA 180

QY 208 LeuSerLeuIleLeuHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 227  
DB 181 CTTTCTTAATAAACAATCCAGAGGTCTCCACATGAGTGTGGGAAAGCCATAGAACAT 240

QY 228 LeuAlaIleGlnGlyAsnArgPheHisPheAspIleLeuValProProLeuHisIleAlaIle 247  
DB 241 TTGGCCAAACAAGAAATAGATTTTCATTTTGACATCAAACTCCCTTGCAATCACTTAA 300

QY 248 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetLeuIle 267  
DB 301 AATGTGATTTTCTTTTACTGACCTTCACACGTTACTGATTAATAATATGAAAAG 360

QY 268 GluIleGluGluGlyIleGluIleGlyGlnIleLeuSerSerAlaAlaAspIleAlaIle 287  
DB 361 GAAAAAAGAGAAAGGTATTTGAAAGGGGCAAAATCCGTCTTGACGACGACATTCCTGCC 420

QY 288 ThrValGlnHisThrMetAlaCysHisLeuValIleValArgThrHisArgAlaIleLeuPhe 307  
DB 421 ACACTACGACACAAATGGCAATGATCTTTGTGAAAGAAACATCATGGGCTATTCGTGTT 480

QY 308 CysIleGluArgAspLeuLeuProGluAsnAsnAlaValIleuValAlaSerGlyGlyVal 327  
DB 481 TGTAGCAGAGAGACTTTGTTACTCAAAATTAATGCACTACTGTTGATCTGGTGTGTC 540

QY 328 AlaSerAsnPheTyrlleArgAlaIleGluIleLeuThrAsnAlaThrGlnCysThr 347  
DB 541 GCAGTAACTTTATATCCGACAGAGCTTGGAAATTTTAACAAACGCAACAGGTGCACT 600

QY 348 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIlePheGly 367  
DB 601 TTGTTGTGTCCTTCCTCCACAGACTATGCACGTGATTAATGCAATTAATGATTCATGAA 660

QY 368 IleGluArgLeuArgAlaGlyIleGluGlyIleLeuHisAspIleGluGlyIleArgGly 387  
DB 661 ATTTGAAAGACTAGCTGCTGGCATTTTACATGACATTAAGAGCATTCGCTATGAA 720

QY 388 ProIleCysProLeuGlyValaSerIleSerIleGluValGlyGluAlaSerIleVal 407  
DB 721 CCAAAATGTCCTCTTGAGATTAAGCATATCAAAAGAAAGTTGGAGAGACTTCATTAAGTA 780

QY 408 ProGluLeuIleuMetGluIle 414  
DB 781 CCACAATTAATAATGAGAGATA 801

RESULT 13  
US-10-649-273-23

; Sequence 23, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; CURRENT FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-649-273-23

Alignment Scores:  
Pred. No.: 1,49e-157 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 65.18% Indels: 0  
DB: 17 Gaps: 0

US-10-649-273-2 (1-414) x US-10-649-273-23 (1-1526)

QY 148 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeuValGluPheProPheLeuVal 167  
DB 1 ATGAGAGCTCATGCACTTAATAATGAGTTGACCAATTAAGTAAATTTCTTTTAGTT 60

QY 168 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 187  
DB 61 CTTTGAATTTCTGGAGGTCACTGTCTGTGGCATTAAGTTCAAGAGTTTCAGATTTTCTG 120

QY 188 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArg 207  
DB 121 CTTCTTGAAAGCTTTTGACATGACACAGGTGACATGCTTGACAAAGTGGCAAGAAGA 180

QY 208 LeuSerLeuIleLeuHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 227



; PRIOR FILING DATE: 2001-04-10  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: Patentin version 3.0  
 ; SEQ ID NO 20  
 ; LENGTH: 14364  
 ; TYPE: DNA  
 ; ORGANISM: homo sapiens  
 ; US-10-067-443-20

## Alignment Scores:

Pred. No.:	7,32e-132	Length:	14364
Score:	1186.50	Matches:	313
Percent Similarity:	32.30%	Conservative:	1
Best Local Similarity:	32.20%	Mismatches:	2
Query Match:	55.84%	Indels:	657
DB:	14	Gaps:	4

US-10-649-273-2 (1-414) x US-10-067-443-20 (1-14364)

QY	74	LystrhrglyglYlIleValProProAlaIaIngInLeuHISargGluuSniIlegInarG	93
DB	10623	AGAACAGGTGGATTTGTTCTTCAGAGCTCAACAGCTTCAAGAGAAATTTCAACA	10682
QY	94	IlleValIngInuAlaLeuSerAlaSerGlyValSerProSerAspLeuSerAlaIleAla	113
DB	10683	ATAGTACAAAGAGCTTTTCGCCAGTGAAGTCTCCAGAGTACCTCCAGCAATTGCA	10742
QY	114	ThrThrIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGln	133
DB	10743	ACTCCATTAACCAAGGAGCTTGAAGCTTGAAGCTGGAAGTGGCTTATCATTTAGCTTACAG	10802
QY	134	LeuValGlyGlnLeuLysPhePheIleProIleHISemeGluAlaHISAlaLeu	153
DB	10803	CTGGTAGAGCAGTTAAAGCATTCATCCCATTCATCATATGAGGCTCATGCACTT	10862
QY	154	ThrIleArgLeuThrAsnLysValGluPheProPheLeuValIleuLeuIleSerGlyGly	173
DB	10863	ACTATTAGGTGACCAATTAAGTAGAATTTCTTTTATTTAGTTCTTTGATTTCTGAGGT	10922
QY	174	HisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeu	193
DB	10923	CACGTGCTGTTGGCATTTAGTTCAAGAGATTTCAGATTTCTCTTCATGGAAGAAGCTTTG	10982
QY	194	AspIleAlaProGlyAspMetLeuAspLysVal	204
DB	10983	GACATGACACCAAGTGAATCTTGACAAAGT-AAATTAAGATTAATTTCTCCATTTCTT	11041
QY	204	-----	204
DB	11042	TTTGTATTGTTGTCATTTCACATTAAGTAGACATATGATGTGCTACCACTTCACTTAA	11101
QY	204	-----	204
DB	11102	TATTTCTGAATTTTATCTTAGTAACGAAAAAATTCACATATGCTGAGAAAAATAGA	11161
QY	204	-----	204
DB	11162	AAGAGTAGTACACAATTTATTAATCTTAGCCTTCTTAATTAATGTAAGAGGTTCAT	11221
QY	204	-----	204
DB	11222	ATCTGTACATTAAGCGTGAATAATGTTGCAATACATGTTATGTAATTTTCCAAATTAAGT	11281
QY	204	-----	204
DB	11282	ATGTGAAGAAGACGTGCTTGTAATTAACATATCTGCAAAAAAGTAAATAAGAGAAATAT	11341
QY	204	-----	204
DB	11342	ATAATGATTAACATTAACACATTAATAAGATGCAATGCAAGAAATTAATCAACAAATTAAT	11401
QY	204	-----	204

DB	11402	TACACCAACAGACAGGTCCCCCCCCACCCCCCTTTGTTTAAATACATACAGAGCTACT	11461
QY	204	-----	204
DB	11462	GCCATATATAGAAAAACAACAAACAACAAACAAACAACTGCTTCCACAGTAAATTA	11521
QY	204	-----	204
DB	11522	TAGAAATATAGACAAAGTTCTTATTAATGACCTTACATTAAGCAAGTATTTGCAACT	11581
QY	204	-----	204
DB	11582	TCAAGCCATTTTCCAACATAGAAAGACAAACATAGACAGGGCAGTATGCTCT	11641
QY	204	-----	204
DB	11642	TATTTGGGTGATCATTAAGAACAGGGTGTCTGCTTACCTGAATATCAGCTATAGTC	11701
QY	204	-----	204
DB	11702	TATATTTGCCAAGTATAGCATGTTTATTCATTCAGGGGTTTTGTTTGTATGTAAT	11761
QY	204	-----	204
DB	11762	TTTCAATTAATTTCTTTCATCTTTTCGTTTCAGATATTAATTTATAGACTTAA	11821
QY	205	-----	205
DB	11822	AATATGTTTCTTGAATAGTGGCAGAAAGACCTTTCTTAATTAACATCCAGAGTGTCC	11881
QY	218	ThrMetSerGlyGlyLysAlaIleGluHISLeuAlaLysGlnGluAsnGlyPheHISPhe	237
DB	11882	ACCATAGTGTGGGAAAGCATAGACATTTTGCCAAACAGAAATTAAGATTCATTT	11941
QY	238	AspIleLysProProLeuHISAlaLysAsnCysAspPheSerPheThrGlyLeuGln	257
DB	11942	GACATCAAACTCCCTGATCATGCTTAATAAATTTGATTTTCTTTTACTGAGCTTCAA	12001
QY	258	HisValThrAspLysIleIleMetLysGlyGluGlu	271
DB	12002	CACGTACATGATTAATAATTAAGAAAAAGAAAGAGATTAATTTCTAATTAAGT	12061
QY	272	-----	272
DB	12062	AAAGTTGAAACAGATAAATATCTTGATGTGCTTAAATTAAGCTGCTCATTTCTGAGG	12121
QY	272	YlIleGlyGlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHISThr	292
DB	12122	TATTTGAAAGGGGCAATCTGTTCTTCAGCAGACATTTGCTGCCAGTACAGCACAC	12181
QY	292	rMetAlaCysHISLeuValLysArgThrHISArgAlaIleLeuPheCysLysGlnArgAs	312
DB	12182	AATGGATGTCATCTTGTGAAGAAACACATGGGGCTATCTGTTTGTAAAGAGAGAGA	12241
QY	312	PleuLeuProGlnAsnAsnAlaValLeu	321
DB	12242	CTTGTACCTCAAAATTAATGACATGCTGTAAGTTTATCTCAATTTTATAGTAATAGTTA	12301
QY	321	-----	321
DB	12302	CACCTTGCAATATGTTACTTTTTCACAAAGACCTTGACCTTGTTTAGATGAACAGAT	12361
QY	321	-----	321
DB	12362	CTTTATGCTTATGCTAGCCCTGACAGTATGAATTAATGACAGATAGAAAGACTAACG	12421
QY	321	-----	321
DB	12422	CCATTTCTGTACTAGITGATAGCTTTATGGACAGCTGTATAGCTTATAGCAGACATA	12481
QY	321	-----	321
DB	12482	AGCTAATTTTGCATCTTCTGTGATTTAAAGAGGCTTACAAATTAAGAAAGTAAT	12541



```
QY 321 ----- 321
Db 12542 GCAGTAAGCTGATCATATTTTAGAAAAATAGGTGATTTCTTCATCTTTGATGAA 12601
QY 321 ----- 321
Db 12602 ATCCCTTGTGTGTTGTTTTTTAATAAGCAGTCAATTAGCAGTGGAGGTGTA 12661
QY 321 ----- 321
Db 12662 TCCACTTTCGTGACATAAGTGTATTAAGTTCGATATCCACTATATGTACCAACC 12721
QY 321 ----- 321
Db 12722 AAAATCCCTTAATGTCCTTAAGACCTTGACAAACATCTGTTAACTGTATCTTAA 12781
QY 321 ----- 321
Db 12782 CTTTATTCATTTAAATAATTAATAAGTGGAAAAATGTTAAATGTAGTAATTCAT 12841
QY 321 ----- 321
Db 12842 AGATGAATTTTACATGATATCAAGAATATTTTTCAGAGTTATGTATGTAATAATGCA 12901
QY 321 ----- 321
Db 12902 CAAAATTAATAAAATTTGAGGGCTTAATAATGTCATATGATGAATTAATTAATA 12961
QY 321 ----- 321
Db 12962 AATATTTAGATGAAGTGTGAGAAAAATATCAAAAATGCTAGTAATGTTGTATGCTA 13021
QY 321 ----- 321
Db 13022 TTAGATTTATTAGTAATTTTCTTCCAAATTTTATTAATACATAGATATGTCATCTGCC 13081
QY 321 ----- 321
Db 13082 CATTAACCATCTCAAAATGGATGTTATTAATGTTAATGCTGATATTTTCTCCAGG 13141
QY 321 ----- 321
Db 13142 TTTAATTAGCAGCTTGTCATATCCATATATAGTATTTGGTTTCTCAATTCCT 13201
QY 322 ----ValAserGlyGlyValAlaSerAsnPhetYrIleArgArgAlaLeuGluIleLe 340
Db 13202 TCAGGTTCATCTGTGTGTGTCGAAGTAATCTTATATCCGACAGCTCTGGAATTTT 13261
QY 340 uThrAsnAlaThrGlnCyethrLeuleuCyAsProProAlaArgLeuCyethrAspAsnG 360
Db 13262 AACAAACCGCAACACAGTCATTTGTGTCTCTCCACAGCTATGCACTGATATATG 13321
QY 360 YlleuettIleAlaTrp----- 365
Db 13322 CATTAATGATTCATGTAAGCCACAGATATACGTCTTCACTCAATACTATGTAATAT 13381
QY 365 ----- 365
Db 13382 TAATTGCATTTTATCATACTAAGCCTTCTTCAGATCTTGAGACTATTGATTTTAT 13441
QY 366 -----AsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuH 379
Db 13442 TTTAATGCTTCTTAATTAAGATGTAATGTAAGAAAGACTAGCTGTGGCATTTTAC 13501
QY 379 IsAspIleGluGlyIleArgTyrGluProIys 389
Db 13502 ATGACATGAAAGGATCCGCTATGAAACCAAG 13533
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:38:06 ; Search time 18.9 Seconds  
(without alignments)  
2107.605 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1 MLILTKTAGVFPKSKRKRV.....DISKEVGSASIKVPLQKMEI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	634	29.8	463	2	probable O-sialoglycoprotein endopeptidase
2	553	26.0	387	2	probable O-sialoglycoprotein endopeptidase
3	548	25.8	365	2	probable O-sialoglycoprotein endopeptidase
4	548	25.8	366	2	probable O-sialoglycoprotein endopeptidase
5	540.5	25.4	344	2	probable O-sialoglycoprotein endopeptidase
6	539.5	25.4	359	2	probable O-sialoglycoprotein endopeptidase
7	535.5	25.2	367	2	probable O-sialoglycoprotein endopeptidase
8	524	24.7	421	2	probable O-sialoglycoprotein endopeptidase
9	516.5	24.3	335	2	probable O-sialoglycoprotein endopeptidase
10	492.5	23.2	341	2	probable O-sialoglycoprotein endopeptidase
11	488.5	23.0	337	2	probable O-sialoglycoprotein endopeptidase
12	488.5	23.0	337	2	probable O-sialoglycoprotein endopeptidase
13	484.5	22.8	337	1	probable O-sialoglycoprotein endopeptidase
14	484.5	22.8	337	2	probable O-sialoglycoprotein endopeptidase
15	483.5	22.8	337	2	probable O-sialoglycoprotein endopeptidase
16	475.5	22.4	325	2	probable O-sialoglycoprotein endopeptidase
17	475.5	22.4	342	2	probable O-sialoglycoprotein endopeptidase
18	469	22.1	340	2	probable O-sialoglycoprotein endopeptidase
19	467.5	22.0	354	2	probable O-sialoglycoprotein endopeptidase
20	465.5	21.9	346	2	probable O-sialoglycoprotein endopeptidase
21	465.5	21.9	354	2	probable O-sialoglycoprotein endopeptidase
22	463.5	21.8	343	2	probable O-sialoglycoprotein endopeptidase
23	460	21.6	336	2	probable O-sialoglycoprotein endopeptidase
24	459.5	21.6	346	2	probable O-sialoglycoprotein endopeptidase
25	457	21.5	327	2	probable O-sialoglycoprotein endopeptidase
26	455	21.4	336	2	probable O-sialoglycoprotein endopeptidase
27	455	21.4	348	2	probable O-sialoglycoprotein endopeptidase
28	451.5	21.2	344	2	probable O-sialoglycoprotein endopeptidase
29	439	20.7	344	2	probable O-sialoglycoprotein endopeptidase

30	439	20.7	346	2	probable O-sialoglycoprotein endopeptidase
31	437.5	20.6	312	2	probable O-sialoglycoprotein endopeptidase
32	431	20.3	448	2	probable O-sialoglycoprotein endopeptidase
33	430	20.2	341	2	probable O-sialoglycoprotein endopeptidase
34	409.5	19.3	338	2	probable O-sialoglycoprotein endopeptidase
35	409	19.2	352	2	probable O-sialoglycoprotein endopeptidase
36	406.5	19.1	344	2	probable O-sialoglycoprotein endopeptidase
37	406.5	19.1	344	2	probable O-sialoglycoprotein endopeptidase
38	405	18.8	374	2	probable O-sialoglycoprotein endopeptidase
39	400.5	18.8	346	2	probable O-sialoglycoprotein endopeptidase
40	400	18.8	336	2	probable O-sialoglycoprotein endopeptidase
41	391.5	18.4	335	2	probable O-sialoglycoprotein endopeptidase
42	379.5	17.9	344	2	probable O-sialoglycoprotein endopeptidase
43	361.5	17.0	407	2	probable O-sialoglycoprotein endopeptidase
44	361	17.0	302	2	probable O-sialoglycoprotein endopeptidase
45	353.5	16.6	315	2	probable O-sialoglycoprotein endopeptidase

#### ALIGNMENTS

RESULT 1	
E84888	probable O-sialoglycoprotein endopeptidase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)	
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Jun-2003	
C:Accession: E84888	
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N. M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; Gues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999	
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.	
A:Reference number: A84420; MIMD:20083487; PMID:10617197	
A:Accession: E84888	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-463 <STO>	
A:Cross-references: GB:AE002093; NID:G2583127; PIDN:AAB82636.1; GSPDB:GN00139	
C:Genetics:	
A:Gene: AtG945270	
A:Map position: 2	
C:Superfamily: O-sialoglycoprotein endopeptidase	
Query Match	29.8%; Score 634; DB 2; Length 463;
Best Local Similarity	37.9%; Pred. No. 2e-43;
Matches 145; Conservative 61; Mismatches 129; Indels 48; Gaps 7;	
QY	38 IVLGIEFTSCDDPRAAAYVDENGAVLGEAHSQTEVHLKTGIVPPAAOQLHRENIQRYOE 97
DB	85 VVGLGISTCDDPRAAAYVSPFNHLSSC---RAELVQYGVAVKQABEASRVYIDRVOD 141
QY	98 ALSASGVSPSDSAIATTTKPGIALSLGVGLSPSLQVQCKKPPPIHMEAHATIRL 157
DB	142 ALDKANLTERKDSAVAVTIGPGISLCLRGVRAARAVAGFSLPIGVHMEHVALVRL 201
QY	158 T-NKVEPPVLIIISGCHLALVQGVSDFLIGKSLDIAPEGMDLVKVARSLIKHPEC 216
DB	202 VQGLISFPFMALLISGCHMLVLAHLQYTGITVDAIGAPKTAACGLDWH--- 258
QY	217 STMSGKALIEHLAKQGNRHPDILKPLHAKNDESPGLQVHTDTIKMKKEBEGIEKG 276
DB	259 -RSGGPAVEELAEEDASVKNFVPMKHKDQNFSAIGLKVRLAIEKK----- 308
QY	277 QILSSNADIAATVQHTMAAGLVTETRAILFCQKORILLPONNAVLVAASGVASNFYRRA 336
DB	309 --IRNRADIASFORAVAVLHEKCRALIDMALE---LEPSIGHMVISGCVASNKTVRLR 363
QY	337 LEILTNATQCTLLCEPPRLCTDNGIMIANNGIERLAGLILADIGIRYE----- 387
DB	364 LNNIVENKRLAKVCPPLSLCTDNGVAVVAVMTGLEHFRVG-----RDPPPPATEPE 413
QY	388 -----PKCPFGVDSKEVGEA 403

Db 414 DYVYDLRPRMPLGEBYAKGRSEA 436

## RESULT 2

E71711

probable o-sialoglycoprotein endopeptidase (gcp) RP037 - Rickettsia prowazekii

C/Species: Rickettsia prowazekii

C/Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 09-Jul-2004

C/Accession: E71711

R/Anderson, S.G.E.; Zomorodipour, A.; Anderson, J.O.; Sicheitiz-Ponten, T.; Alsmark, U. Nature 396, 133-140, 1998

A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A/Reference number: A71630; MUID:99039499; PMID:9823893

A/Status: preliminary

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-387 <AND>

A/Cross-references: UNIPROT:Q9ZEA8; GB:AJ235270; GB:AJ235269; NID:G3860572; PIDN:CAA1450

A/Experimental source: strain Madrid R

C/Genetics:

A/Gene: gcp; RP037

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 26.0%; Score 553; DB 2; Length 387;

Best Local Similarity 35.1%; Pred. No. 5.6e-37;

Matches 133; Conservative 56; Mismatches 130; Indels 60; Gaps 6;

QY 37 KIVIGIETSCDDTAAYVDEFGNVLGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRIYO 96

Db 2 KIIIGIESSCDDTAISITERRKILSNIIISQNTHEAVFGVVPRIARSHLSNLDQALK 61

QY 97 EALSASGVSPDLSAIAITIKPGLALSGLVGSFSLQVGLKPPPIPHHMEAHALT 156

Db 62 NVLKKSNTLEISIAITSGPGLIGYVSGMFAKSPASSAKKPPAIINHLBGHALTKR 121

QY 157 LTNKVEPFLVLLISGHCILALVOGVSDPFLILGKSLDAPGMDLVKARRLSLTKHPEC 216

Db 122 LTNMISVYLLILSLSGHCQGFPAVGLGKRYKILGTTIDDAVETIDKAKAMTL----- 175

QY 217 STMSGGAIEHLAKQGNRFHDIKPLHAKNCDSPFTGLQHTVDKIIMK-KEKEGIERK 275

Db 176 -SFGGPEIEKRAKGNPHKYPKPIINSNGCNMSFSGLKTAVRVLINWLKEVNDV-- 232

QY 276 GQISSADIAATVQHTMACHLVKTRHALIFCKQ-----RD-- 312

Db 233 -----INDIASFQITGAILISSKQDAIRLYKQILNDYEDINHPPTKLNKSPKDEP 286

QY 313 -----LLPON-----NAVIVASGVASNFYIRBALEILTNATQCTLLCP 351

Db 287 NWKPLECTTRKRYRHIQNSRNSNLNDITIVAGVAAANKVLOEILSDCTREYGRLLAP 346

QY 352 PPRLCDNGIMIANWGIR 370

Db 347 PMHLCTDMAAMIAVAGLER 365

RESULT 3

AB2902

O-sialoglycoprotein endopeptidase gcp [imported] - Agrobacterium tumefaciens (strain C58

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C/Accession: AB2902

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, L.

Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCLell

; Karp, P.; Romero, P.; Zhang, S. Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, B.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; MUID:21608550; PMID:11743193

A/Accession: AB2902

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-365 <KUR>

A/Cross-references: GB:AE008688; PIDN:AAI43632.1; PID:g17741154; GSPDB:GN00186

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Gene: gcp

A/Map position: circular chromosome

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 25.8%; Score 548; DB 2; Length 365;

Best Local Similarity 36.9%; Pred. No. 1.3e-36;

Matches 137; Conservative 58; Mismatches 142; Indels 34; Gaps 8;

QY 39 VLGIETSCDDTAAYV----DETGNVLGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRI 94

Db 7 ILGIETSCDETAISIVRRADRGELVSDVLSQLEHSAVGGVPEIARAVAEALDTL 66

QY 95 VOBALSASGVSPDLSAIAITIKPGLALSGLVGSFSLQVGLKPPPIPHHMEAHALT 154

Db 67 VESALDQAGVKLADVDAIATSGPGLIGLVLGLMTGKAIKAAGKPLVAINHLBGHALT 126

QY 155 IRLTNKVEPFLVLLISGHCILALVOGVSDPFLILGKSLDAPGMDLVKARRLSLTKH 214

Db 127 ARLTDGLSPPYLMLLVSGHTQLVLRGVGEYERWGTITDDALGEAPDKAKLGL-PPY 185

QY 215 ECGTMSGGAIEHLAKQGNRFHDIKPLHAKNCDSPFTGL-----QHTVDKIIMKEKE 270

Db 186 -----GEPAVENAAKQDPDRPLPRPMVGEARLDFSGGLKTAVRQAATALAPLSEQD 239

QY 271 EGIEKQIISADIAATVQHTMACHLVKTRHALIFCKQRLDPONNA--VLVASGVA 328

Db 240 -----IADICASFQKAVSRILKDRIGLARFKVE--FPHINGEPALVAVGVA 286

QY 329 SNFYIRBALEILTNATQCTLLCPPRLCDNGIMIANWGIRLAKGILHIDIGIRYEP 388

Db 287 ANQELRQTLQALCDTGFHFVAPPHRLCTDMAAMIAWAGLERMAEG---ROADALEVAP 342

QY 389 KCPILGVIDISKE 399

Db 343 RSRWPLDGSAB 353

RESULT 4

D97677

probable o-sialoglycoprotein endopeptidase (glycoproteinase) [imported] - Agrobacterium t

C/Species: Agrobacterium tumefaciens

C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002

C/Accession: D97677

R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Bianchard, M.; Gurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume

A/Reference number: A97359; MUID:21608551; PMID:11743194

A/Accession: D97677

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-366 <KUR>

A/Cross-references: GB:AE007869; PIDN:AAK8373.1; PID:g15157858; GSPDB:GN00169

A/Genes: AGR C 4806

A/Map position: circular chromosome

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 25.8%; Score 548; DB 2; Length 366;

Best Local Similarity 36.9%; Pred. No. 1.3e-36;

Matches 137; Conservative 58; Mismatches 142; Indels 34; Gaps 8;

QY 39 VLGIETSCDDTAAYV----DETGNVLGEAHSQTEVHLKTGGIVPPAAOQLHRENIQRI 94

Db 8 ILGIETSCDETAISIVRRADRGELVSDVLSQLEHSAVGGVPEIARAVAEALDTL 67

QY 95 VOBALSASGVSPDLSAIAITIKPGLALSGLVGSFSLQVGLKPPPIPHHMEAHALT 154

Db 68 VESALDQAGVKLADVDAIATSGPGLIGLVLGLMTGKAIKAAGKPLVAINHLBGHALT 127



Matches 136; Conservative 54; Mismatches 144; Indels 25; Gaps 7;

QY 38 IVLGIEISCDPTAAAVV-----DEGNVIGEAHISQTEVHLKTGIVPPAAOQLHRENIOR 93  
 Db 10 IILGIEISCDPTAAVSVRRADGIVTVLSSVIGTFPERKAPFGGVVPEIAAAHVESIDA 69  
 QY 94 IVOEALSASGVSPDLSAATTIKPGLALSIGVGSFSLQVLGOLKKPPIPIHMEAHAL 153  
 Db 70 IAAEAVRAGVFGDLDDVAATAGPGLVGVVWGLAFGKAVALLAGAPLVANHLBGHAV 129  
 QY 154 TIRLTNNVPPPLVLLISGHCILALVQGVSDPFLILGSLDIAPIADMIDKVARSLIKH 213  
 Db 130 SARLGADIAVYFPLILVSGHCQLLEVSGVCKKLGITIDDAAGAEAFDKIKSLGL-PE 188  
 QY 214 PECSMGSKAIEHLAKGNRHFPIKPRPLHAKKCDSPFTGLQVNTDKIIMKKKEBEI 273  
 Db 189 P-----GGPALBEKLVAGSDPFRVALPRALGKRCDCSPSGLKTAAAIATLTDD-- 240  
 QY 274 EKGQILSSAADIATVQHTMACHLVKRTTRAILPCKORDLIPONNAVLYVSGVASNFYI 333  
 Db 241 -----ARRDLAAGVQAAIARQLSERVDRAMKLYK--DSHPEDLRFVAVAGVANAAGV 291  
 QY 334 RRALEILTNAQCTLLCPPEPLCTDNGIMIANNGIERLAGLIIHDIIGIRYEPKCP 392  
 Db 292 RAALLADCEKNGFSFAPAPPLACTDNAAIMIALAGERL--ALGIFDLDALA-RPRWPL 347

## RESULT 8

T18825

hypothetical protein C01G10.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000

C:Accession: T18825

R:Matthews, L.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19027

A:Accession: T18825

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-421 <WIL>

A:Cross-references: EMBL:Z81030; PIDN:CA802716.1; GSPDB:GN00023; CESP:C01G10.10

A:Experimental source: clone C01G10

C:Genetics:

A:Gene: CESP:C01G10.10

A:Map position: 5

A:Initrions: 31/2; 122/2; 177/1; 272/3; 315/1; 353/2

C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 24.7%; Score 524; DB 2; Length 421;

Best Local Similarity 33.4%; Pred. No. 1,4e-34;

Matches 130; Conservative 77; Mismatches 146; Indels 36; Gaps 8;

QY 39 VIIGIETSCDPTAAAVVDETNVIGEAHISQTEVHLKTGIVPPAAOQLHRENIORIVQEA 98  
 Db 25 VIIGIETSCDPTAAVAVNEKREILSSBRTERRAIQKQGSINPVCALQREHLPRILEKC 84  
 QY 99 LSASGVSPDLSAATTKIGLALSIGVGSFSLQVLGOLKKPPIPIHMEAHALTRILT 158  
 Db 85 LNDAGTSBKDLDAVAVTVTPGLVIALKEGISAIFAKKGRLLPLIPVHMRALHALLILV 144  
 QY 159 -NKVEFPLVLLISGHCILALVQGVSDPFLILGSLDIAPIADMIDKVARSLIKHPECS 217  
 Db 145 DDSVRFPPSAVLLISGHALISVADEVKFKLVGVSQSGPGECDIKVAQQLDGL-GSEFD 203  
 QY 218 TMSGGKALEHLAKOQN---RHFPIKPRPLHAKNCDSPFTGLQVNTDKIIMKKKEBEI 274  
 Db 204 GIHVGAIVEILASRSASGCHLRYPFLPVPVKANANFOIKSYLMLERLAKNSTSID 263  
 QY 275 KGQILSSAADIATVQHTMACHLVKRTTRAILPCKORDLIPONNAVLYVSGVASNFYIR 334  
 Db 264 -----IPDFCASLQNTVARIHSSKMLHPFESLSBQELPFGQ---LVIGGVAAANQYIF 313  
 QY 335 RALEILTNAQCTLLCPPEPLCTDNGIMIANNGIERLAGLIIHDIIGIRYEP----- 388

Db 314 GAISKLSAAHNVTTIKVLLSLCTDNAMIAVYSGL-----LMLVNRSEAIWPRPNDIPDR 367

QY 389 -----KCPFGVIDISKEVGASIKVPQLKM 412

Db 368 IYAHARSDIGTDASSSEI---IDTPRRKL 392

## RESULT 9

G70369

sialoglycoproteinase - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004

C:Accession: G70369

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove

V. Nature 392, 353-358, 1998

A>Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: G70369

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-335 <AQF>

A:Cross-references: UNIPROT:O66986; GB:AE000708; NID:G2983356; PIDN:AA006951.1; PID:G2983

A:Experimental source: strain VFS

C:Genetics:

A:Gene: gcp

C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 24.3%; Score 516.5; DB 2; Length 335;

Best Local Similarity 36.5%; Pred. No. 4.2e-34;

Matches 130; Conservative 66; Mismatches 125; Indels 35; Gaps 10;

QY 40 IGIETSCDPTAAAVVDETNVIGEAHISQTEVHLKTGIVPPAAOQLHRENIORIVQEA 99  
 Db 4 LAVETSCDETALAIYDDQKGVIGNVILSQAVVHSPGVVPELSAREHTRNIIPIFDRL 63  
 QY 100 SASGVSPDLSAATTKIGLALSIGVGSFSLQVLGOLKKPPIPIHMEAHALTRILT 159  
 Db 64 KESRINLEEDIDPISFLTLEGLILSLVGAFAALAEYERKPLVPVPHHLEGHVYFLEK 123  
 QY 160 KYEPFLVLLISGHCILALVQGVSDPFLILGSLDIAPIADMIDKVARSLIKHPECS 219  
 Db 124 KYEPFLVLLISGHCILALVQGVSDPFLILGSLDIAPIADMIDKVARSLIKHPECS 219  
 QY 220 SGGKALEHLAKOQNRFHPIKPRPLHAKNCDSPFTGLQVNTDKIIMKKKEBEIEGQIL 279  
 Db 178 -GGPIIDRLAKGKGL-YPLPKPLMEEGNINFSFGLK--TALLMLKKKEKVRK---- 228  
 QY 280 SSAADIATVQHTMACHLVKRTTRAILPCKORDLIPONNAVLYVSGVASNFYIRALEI 339  
 Db 229 ---EDIAVSFOETVAILLEKS---LWMLKKTGIR---LVVGVGSANSRLR--EV 274  
 QY 340 LTNATQ---CTLLCPPEPLCTDNGIMIANNGIERLAGLIIHDIIGIRYEPKCP 392  
 Db 275 FKXASQYGFELTYIPPSLSTDNAMIAVAGMERFRGVAVPLDVP---QPMIPL 327

## RESULT 10

H83572

O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: H83572

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brj

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: H83572

A:Status: preliminary

A:Molecule type: DNA

A;Residues: 1-341 <STO>  
A;Cross-references: GB:AE004494; GB:AB004091; NID:g9946446; PIDN:AAG03969.1; GSPDB:GN001  
A;Experimental source: strain PAOI  
C;Genetics:  
A;Gene: gcp; PA0580  
C;Superfamily: O-sialoglycoprotein endopeptidase

Query Match 23.2%; Score 492.5; DB 2; Length 341;  
Best Local Similarity 36.1%; Pred. No. 3.7e-32;  
Matches 126; Conservative 61; Mismatches 143; Indels 23; Gaps 9;

QY 39 VLGIETSCDDTAAAVDETGAVLGEAHSQTEVHLKGTGIYPPAAQOLHRENIQRIYGEA 98  
DB 3 VLGIETSCDDTAAAVDETGAVLGEAHSQTEVHLKGTGIYPPAAQOLHRENIQRIYGEA 62  
QY 99 LSASGVSPPDLSAATITIKPGLALSLGVLSQVLSQVQKKPFIHMHMAHLTIL 158  
DB 63 LDESGCTPADIDALAYTAGPGLVGLAVGASCAQMAFPAVGVHMEGHLLAPMLE 122  
QY 159 NK-VEPFLVLLISGHCLLALVOGSDPFLLGKSLDIAPGMDLKVARRLSLKHPECS 217  
DB 123 EQPRFPFVALLVSGHQLVAVDGIKRYQLGESVDDAGAFDKTALIGL-GYP--- 178  
QY 218 TMSGKAI EHLAKQGNRFPDIKPLHAKNCDPFSFTGLQHTVDKIIMKKEKEGIEKQ 277  
DB 179 ---GGPEIARLAERGTGPFVFPFPRMTDRPGLDPSFGIKTFPTLN-TWQRCVEAGDDSEQ 234  
QY 278 ILSSADIAATVOHTMACHLVKTRHAILFCQKRDLLPQNNAVLVASGVASNYRRL 337  
DB 235 ---TRCIAALFQAVAVETLLIKCRAL---KQTEL--KN--LVIAAGVSANQALRSGL 283  
QY 338 EILTNATQCTLLCPRLCTDNGIMIANWGIERLHAGLILHDIGIYEPKCP 392  
DB 284 EKMLGEMKGVFYARPFCTDNGAMTAYAGCQRLIAG--QHDGPALISVQPRWPM 335

## RESULT 11

C91122  
probable O-sialoglycoprotein endopeptidase [imported] - Escherichia coli (strain O157:H7  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C;Accession: C91122

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: C91122  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-337 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BA037370.1; PID:G13363420; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RMD 0509952  
C;Genetics:  
A;Gene: EC63947  
C;Superfamily: O-sialoglycoprotein endopeptidase

Query Match 23.0%; Score 488.5; DB 2; Length 337;  
Best Local Similarity 35.2%; Pred. No. 7.8e-32;  
Matches 125; Conservative 55; Mismatches 146; Indels 27; Gaps 9;

QY 39 VLGIETSCDDTAAAVDETGAVLGEAHSQTEVHLKGTGIYPPAAQOLHRENIQRIYGEA 98  
DB 3 VLGIETSCDDTAAAVDETGAVLGEAHSQTEVHLKGTGIYPPAAQOLHRENIQRIYGEA 62  
QY 99 LSASGVSPPDLSAATITIKPGLALSLGVLSQVLSQVQKKPFIHMHMAHLTIL 157  
DB 63 LKESGTLAKOIDAAYTAGPGLVGLAVGATVGSGLAPAMVPAIVHMHGHLAPMLE 122  
QY 158 TNKVEPFLVLLISGHCLLALVOGSDPFLLGKSLDIAPGMDLKVARRLSLKHPECS 217  
DB 123 DNPEPFPVALLVSGHQLVAVDGIKRYQLGESVDDAGAFDKTALIGL-DYP--- 178

QY 218 TMSGKAI EHLAKQGNRFPDIKPLHAKNCDPFSFTGLQHTVDKIIMKKEKEGIEKQ 277  
DB 179 ---GGPLSLKMAAQTAGFVFPFPRMTDRPGLDPSFGIKTFPANTL-----RNGTDD-- 229  
QY 278 ILSSADIAATVOHTMACHLVKTRHAILFCQKRDLLPQNNAVLVASGVASNYRRL 337  
DB 230 ---QTRADIAFAFEDAVDTLMICKRAL-----DLTGFR--LVMAAGVSANRTIRAKL 279  
QY 338 EILTNATQCTLLCPRLCTDNGIMIANWGIERLHAGLILHDIGIYEPKCP 392  
DB 280 AEMMKRKGSEVYARPFCTDNGAMTAYAGMVFKA--GATADL-GVSVPFPMPL 331

## RESULT 12

B85967  
probable O-sialoglycoprotein endopeptidase y9jd [imported] - Escherichia coli (strain O1:  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C;Accession: B85967

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousle, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: B85967  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-337 <STO>  
A;Cross-references: GB:AE005174; NID:g12517643; PIDN:AAG58198.1; GSPDB:GN00145; UMGP:2443  
A;Experimental source: strain O157:H7, substrain EDL533  
C;Genetics:  
A;Gene: y9jd  
C;Superfamily: O-sialoglycoprotein endopeptidase

Query Match 23.0%; Score 488.5; DB 2; Length 337;  
Best Local Similarity 35.2%; Pred. No. 7.8e-32;  
Matches 125; Conservative 55; Mismatches 146; Indels 27; Gaps 9;

QY 39 VLGIETSCDDTAAAVDETGAVLGEAHSQTEVHLKGTGIYPPAAQOLHRENIQRIYGEA 98  
DB 3 VLGIETSCDDTAAAVDETGAVLGEAHSQTEVHLKGTGIYPPAAQOLHRENIQRIYGEA 62  
QY 99 LSASGVSPPDLSAATITIKPGLALSLGVLSQVLSQVQKKPFIHMHMAHLTIL 157  
DB 63 LKESGTLAKOIDAAYTAGPGLVGLAVGATVGSGLAPAMVPAIVHMHGHLAPMLE 122  
QY 158 TNKVEPFLVLLISGHCLLALVOGSDPFLLGKSLDIAPGMDLKVARRLSLKHPECS 217  
DB 123 DNPEPFPVALLVSGHQLVAVDGIKRYQLGESVDDAGAFDKTALIGL-DYP--- 178  
QY 218 TMSGKAI EHLAKQGNRFPDIKPLHAKNCDPFSFTGLQHTVDKIIMKKEKEGIEKQ 277  
DB 179 ---GGPLSLKMAAQTAGFVFPFPRMTDRPGLDPSFGIKTFPANTL-----RNGTDD-- 229  
QY 278 ILSSADIAATVOHTMACHLVKTRHAILFCQKRDLLPQNNAVLVASGVASNYRRL 337  
DB 230 ---QTRADIAFAFEDAVDTLMICKRAL-----DLTGFR--LVMAAGVSANRTIRAKL 279  
QY 338 EILTNATQCTLLCPRLCTDNGIMIANWGIERLHAGLILHDIGIYEPKCP 392  
DB 280 AEMMKRKGSEVYARPFCTDNGAMTAYAGMVFKA--GATADL-GVSVPFPMPL 331

## RESULT 13

O06CR6  
O-sialoglycoprotein endopeptidase (BC 3.4.24.57) - Escherichia coli (strain K-12)

C;Species: Escherichia coli  
C;Date: 30-Jun-1988 #sequence\_revision 17-Oct-1997 #text\_change 09-Jul-2004  
C;Accession: F55024; 525025  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F55094  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-337 <BLAT>  
A:Cross-references: UNIPROT:P05852; GB:AEO00388; GB:U00096; NID:g1789441; PIDN:AAC76100.  
A:Experimental source: strain K-12, substrain MG1655  
R:Nesin, M.; Iupski, J.R.; Svec, P.; Godson, G.N.  
Gene 51, 149-161, 1987  
A>Title: Possible new genes as revealed by molecular analysis of a 5-kb Escherichia coli  
A:Reference number: A91573; MUID:87248073; PMID:3297921  
A:Accession: D29049  
A:Molecule type: DNA  
A:Residues: 1-135,'C',137-337 <NES>  
A:Cross-references: GB:M16194; GB:X00773; NID:g147764; PIDN:AA72575.1; PID:9551834  
C:Genetics:  
A:Gene: ygiJ  
A:Map position: 67 min  
A:Superfamily: O-sialoglycoprotein endopeptidase  
C:Keywords: hydrolase; metalloproteinase; zinc  
F:111,115/Binding site: zinc (His) #status predicted

Query Match            22.8%; Score 484.5; DB 1; Length 337;  
Best Local Similarity 35.1%; Pred. No. 1.6e-31;  
Matches 125; Conservative 54; Mismatches 146; Indels 29; Gaps 9;

OY        39 VLGIEIETSCDDTAAVAVDEGTNGVLGEAIHSQTEVHLTKTGIVPPAAOQLIRENIORIVOE 98  
         |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db        3 VLGIEITSCEGTGAIRYDDEKGLLANOLYSQVKLHADYGGVPELASRDHVRKTVPILIQ 62  
         |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
OY        99 LSAGVSPEBDSLAIATTIKGGLAISLVGSLFSLOLVQCLKPFPPIHMEAHALTIRL- 157  
         ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db        63 LKESGITLAADIDAVAYTAGPLVGALVLGAATVGRSLAPAMDPAIPVHMEGILLAPMLE 122  
         ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
OY        158 TNKVPEPFVLIISGGCILATVOGVSDLLIGSLDIAPGMLDKVARRLSIKHPES 217  
         |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db        123 DNPEPFPALVALSGGHTQLISVTGICGYELLGESIDPAAGAFAFDATCLGL-DYP--- 178  
         |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
OY        218 TWSGKAIAHLAKQGNRFHFEDIKRPFLHNAKNCDSEFTGLOHTDTKIIMKKKEEGIEKGQ 277  
         ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db        179 ---GGELISKMAAQAGTAGRVFPRPMIDRPGLDPSFGSLKTFPAANTI-----KDNGIDD-- 229  
         ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
OY        278 ILISSAADIAATVOHTMACHLVKRTTRAILFCQKQRDLPONN-AVIVASGVVASNFYIRRA 336  
         |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db        230 --QT RADIRARAFEDAVVDLMICKRA-----LDQTGFKRLVMAGSVSANRTLRAK 278  
         |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
OY        337 LEILTNATQCTLLCPPRRLCTDNCGIMIAMNGIERLRAGLGIHADI EGIIYEPRCP 392  
         |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db        279 LAEMMKRRGGEVYARPEFCTDNGAMTAAYAMVERKA--GATADL-GSVVRPRWPL 331

	Query Match	22.8%	Score 484.5	DB 2	Length 337
	Best Local Similarity	34.6%	Pred. No. 1.6e-31		
	Matches 123	Conservative	55	Mismatches 150	Indels 27
					Gaps 8
Qy	39	VLGIEIISCDPIPAAVDETVNGVIGEAHISQTEVHLKTGGIVPRAQOLHRENIQRIVOEA	98		
Db	3	VLGIETSCDEGIAIYDDKKGLLANQYISYKGLHADYGGVPELASRDIHVKTVEPLIOAA	62		
Qy	99	LSASGVSSDLSAIAATTTIKPGLTALSIGVGSFSLIQVQOLKKPPIPIHMEHAHTIIL	157		
Db	63	LKEAATLASDIDAAYTAAPGLVGAALLVGATVGSLSAPAMNVPAIVAHMEGHILAPMLE	122		
Qy	158	TNKAEPFLVLLISGGHCLTALVOGVSPDFLLIGSLDPAFGDMDLKVARRSLIKHPECS	217		
Db	123	DNPPDFPFPVALLVSGGHQQLISVGIQGYELIGESSIDPAAGEAPDKTAKLGL-DYF---	178		
Qy	218	TMSGGKAI EHLAKQGNREHPDIKPEPLHAAKCNDFSFTGLQHVTDKI INKKEKEGIEKGQ	277		
Db	179	---GGPMLSKMAASQGTAGRFVPRPMTDRPGIDPSFGLKTFEANTIRISNGDE-----	229		
Qy	278	ILSSADIDIAIVQHTMACHLVKYKTHRAILPFCKORDLPQNNAAVLVASGVSANFTIRAL	337		
Db	230	--QTRADIARAFEBADAVDTL-----MIKK-PALBESTGFKRMVAGGVSSANRTLRKTL	279		
Qy	338	EILTNAQOCTLCPPRLCTDNGIMIAMNGIERIRAGIGLIHDIIGIRYBPICPL	392		
Db	280	AEHQQKRGGEVFTYARPECFIDNGAMIAVAGVRFKA--GVTAADL-GVTVYRPRMPL	331		

RESULT 15  
A10079  
probable glycoproteinase gcp [imported] - *Yersinia pestis* (strain CO92)  
C:Species: *Yersinia pestis*  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: A10079  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
demo-Tarrag, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; F  
ll, M.; Kuterford, K.; Simmons, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, I  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11585630  
A:Accession: A10079  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <KUR>  
A:Cross-references: GB:AL590842; FIDN:CAK89500.1; PID:gl5978736; GSPDB:GN00175  
C:Genetics:  
A:Name: gcp  
i:Superfamily: O-sialoglycoprotein endopeptidase

RESULT 14  
AG0892  
Probable glycoprotease [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhi (S  
C/Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A/Note: this species has also been called *Salmonella typhi*  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C/Accession: AG0892  
R.Parkhill, J., Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher  
th, T.; Connerton, S.; Croxall, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar  
A/Reference number: AB0502; KMDID:21534947; PMID:11677608  
A/Accession: AG0892  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-337 <PAR>  
A/Cross-references: GB:AL513382; PIDN:CAD0773.1; PID:gl6504285; GSPDB:GN00176  
C/Genetics:  
A/Gene: STY3387  
A/Superfamily: O-sialoglycoprotein endopeptidase

	Query Match	22.8%	Score 483.5;	DB 2;	Length 337;	
	Match Local Similarity	34.6%;	Pred. No. 2e-31;			
	Batches 123;	Conservative	54;	Mismatches 150;	Indels 29;	Gaps 7
Oy	VLAGIETSCDDPAAAVVDEGTENVLGEBAHSQCEVHLTKGIVPPAQAOLHRENIORIOVEA					98
Db	VLGIETSCDEDEIGIAVYDDKAGLIANQLQTSQYKLAHDVGGVPELLASRDRHVKTVPVLQDA					62
Oy	LSAGVSPSDLSAIATTIKPGALASLGVSFSFIQVLQKKPFPIHHMEHAALTIR-					157
Db	LKEANILSKDIDAAVAATPGELVGNALLGATIGRALAFAMGWPAVPAVHMHECHLTAPMLE					122
Oy	TNKVBPFLVYLISGGHCILLNVGVSPDFLLSGSLDIAPGDMDLKVARRLSIHQECS					217
Db	ENABEPFPVALLVSGGHQTLSVTGIGBYLLIGSVVDVDAAGEAPDKTAKTLGL-DYP---					178
Oy	TMSGGALEHTLAKOGNRFHPDIKPLHAKKNCDSPFTGLQHVTDXIIKKKEBEGIEKQO					277
Db	---GGPMLSRAAQGTGVRFTFPPRPMTDRPGLDIFSGLKTPFAANTITIANGD-----					229
Oy	ILSSADIDIAATVOHTMACHLYKTRTHRALIFCKQRDLDPNN-AVLVASGCVASFYIRRA					336

Db 230 --OTRADIARAFEDAVVDTLAKSKRA-----LDOTGPRCLVIAGGSANOTLRK 278  
Qy 337 IEILTNATQCTLLCPPEPRCTDNGIMIANGTIERLAGHGLHDIGIRYEPKPL 392  
Db 279 LADMMQKRGGEVFTYARPEPCTDNGAMTAYAGWRLRSN--LNSELTSVSVPRWPL 331

Search completed: February 16, 2005, 13:09:20  
Job time : 19.9 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:19:07 ; Search time 4918.95 Seconds  
(without alignments)  
3203.653 Million cell updates/sec

Title: US-10-649-273-2  
Perfect score: 2125  
Sequence: 1 MLILTKTAGVFFKPSKRVY.....DISKEVGASIKVPLQKMEI 414

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+g2n.model -DEV=x1h  
-Q=/cgn2\_1/USPTO.spool/US10649273/runat\_14022005\_114703\_16411/app\_query.fasta\_1.1429  
-DB=EST -QPMT=fastcap -SUFFIX=rcs -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=b1ts -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273.QCGN\_1\_1\_6799.Qrunat\_14022005\_114703\_16411 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEDUPLY -NEG SCORES=0 -WAIT -DSPBLCK=100 -LONGLOC  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hc.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1827	86.0	2284	3	AK045669 Mus muscu
2	1819	85.6	1622	3	AK011265 Mus muscu
3	1634	76.9	1609	3	BC030671 Mus muscu
4	1447	68.1	870	5	BQ423651 AGENCOURT
5	1353.5	63.7	852	5	BX391919 BX391919
6	1279	60.2	922	5	BQ961028 AGENCOURT
7	1173	55.2	701	2	BR740611 601595739
8	1158	54.5	749	7	CK982692 4115331 B
9	1144	53.8	800	7	CO738006 SLMH03C19

10	1132	53.3	1109	5	BM907988
11	1111	52.3	765	7	CF114247
12	1097	51.6	640	5	BQ636028
13	1045	49.2	730	7	CN823245
14	1038.5	48.9	661	7	CK833139
15	1018.5	47.9	879	5	BQ433135
16	1017	47.9	1173	6	CD508917
17	997	46.9	723	7	CK942291
18	978	46.0	579	7	CK819035
19	966	45.5	822	7	CF257246
20	963	45.3	577	7	CV028547
21	954.5	44.9	808	7	CK792978
22	954	44.9	833	5	BU159877
23	951.5	44.8	728	4	BG436856
24	935.5	44.0	879	5	BU256052
25	935	44.0	871	2	BF308169
26	925	43.5	658	7	CF362328
27	923	43.4	811	7	CN232675
28	920	43.3	859	5	BU292595
29	916	43.1	831	7	CK455183
30	915	43.1	812	5	BU246489
31	912	42.9	878	7	CK456228
32	911	42.9	661	7	CF787040
33	907	42.7	717	7	CK459683
34	906	42.6	597	6	CB272391
35	902	42.4	533	4	BM126453
36	893	42.0	574	5	BQ130908
37	893	42.0	687	7	CF360919
38	879	41.4	737	7	CV073257
39	867	40.8	616	6	CB438468
40	857.5	40.4	789	5	BU242187
41	838	39.4	637	7	CK941819
42	831	39.1	853	7	CR444993
43	825.5	38.6	806	4	BG333362
44	819.5	38.6	792	5	BU403563
45	819	38.5	668	5	BU285782

## ALIGNMENTS

RESULT 1  
AK045669  
LOCUS  
DEFINITION  
Mus musculus adult male corpora quadrigemina cDNA, RIKEN  
Full-length enriched library, clone:B230219017 Product:similar to  
PUTATIVE SIALOGLYCOPROTEASE TYPE 2 (Homo sapiens), full insert  
sequence.

ACCESSION  
AK045669.1 GI:26337528  
VERSION  
AK045669.1  
KEYWORDS  
HTC, CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus

REFERENCE  
AUTHORS  
1 Carninci, P. and Hayashizaki, Y.  
TITLE  
High-efficiency full-length cDNA cloning  
JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE  
99279253  
PUBMED  
10349636  
REFERENCE  
AUTHORS  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL  
Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE  
20493374  
PUBMED  
11042159

REFERENCE  
AUTHORS  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Mishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS
Pujiwke, S., Inoue, K., Togawa, Y., Irawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayaishi, Y.					
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer					
Genome Res. 10 (11), 1757-1771 (2000)					
20530913					
11076861					
4					
The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.					
Functional annotation of a full-length mouse cDNA collection					
Nature 409, 685-690 (2001)					
5					
The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.					
Analysis of mouse transcriptome based on functional annotation of 60,770 full-length cDNAs					
Nature 420, 563-573 (2002)					
6 (bases 1 to 2284)					
Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayaishi, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imomani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takai-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayaishi, Y.					
Submitted (16-JUL-2001) Yoshihide Hayaishi, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)					
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.					
Please visit our web site for further details.					
URL: http://genome.gsc.riken.jp/					
URL: http://lcntom.gsc.riken.jp/					
Location/Qualifiers					
1. .2284					
/organism="Mus musculus"					
/mol_type="mRNA"					
/strain="C57BL/6J"					
/db_xref="PANTOM,DB:B230219017"					
/db_xref="taxon:10090"					
/clone="B230219017"					
/sex="male"					
/tissue_type="corpora quadrigemina"					
/clone_lib="RIKEN full-length enriched mouse cDNA library"					
/dev_stage="adult"					
221. .1465					
/note="unnamed protein product; putative similar to PUTATIVE SIALOGLYCOPROTEASE TYPE 2 (Homo sapiens) (SPTR Q9H4B0, evidence: PASTY, 80%ID, 100length, match=1242)"					
/codon_start=1					
/protein_id="BAC32450.1"					
/db_xref="GI:26337529"					
/translacion="MLMRTAGAIIPKPKSKVGLRFRFSVHPRTLSCHLVAGIEFSCDDTGAAYVDENGVNGLGALHSOTVAGKGVVAPVAAQLENIORLYEFLSACRITPSDLSAIAITTKPGSLALSLGGLSFDLVNQFKKPPPIHMEAHALITLTKRVSFPPVLLISGGHCLATLVGVSDFLLGLSDIADGMDLKVARSLLIKPCESTVMSGKLEIOLAKGNRPHPTINPMONAKCDFEFTGJHITDILTHKEKEBQIEKGK					

[illegible]

[illegible]

TITLE	JOURNAL	REFERENCE	AUTHORS
FUNCTIONAL annotation of a full-length mouse cDNA collection	Nature 409, 685-690 (2001)	5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	Nature 420, 563-573 (2002)	6 (bases 1 to 1622)	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirao, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Iwasa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Komori, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishik, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, P., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submision	Submitted (10-JUL-2000)		Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.jp/) for further details.			
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGGAGGATCCAGACCTCTTTTCTTTTCTTTTAA 3'], cDNA was prepared by using triethanolamine thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGGAGATTCGAGTTAATTAATTAATGCCCCCCCC 3']. cDNA was cleaved with XhoI and SctI. Cloning sites, 5' end: XhoI, 3' end: SctI. Host: SOLR.			
Location/qualifiers			
1. 1622			
/organism="Mus musculus"			
/mol_type="mRNA"			
/strain="CS7BL/6J"			
/db_xref="PANTOM,DB:2610001M19"			
/db_xref="taxon:10090"			
/clone="2610001M19"			
/tissue_type="whole body"			
/clone_lib="RIKEN full-length enriched mouse cDNA library"			
/dev_stage="10 days embryo"			
207. 1451			
/note="unnamed protein product; putative similar to PUTATIVE SIMALOGYPROTEIN TYPE 2 (Homo sapiens) (SPTR Q9H4B0, evidence: PASTY, 80%ID, 100%length, match=1242)"			
/cdon_start=1			
/protein_id="BAB27506.1"			
/db_xref="GI:12847276"			
/translation="MLMRTTAGAIPKPKSKYVGLRPSFVHPRTLSCHLVIGIETFCSCDQAAVVDENGVNGLALHSOTVHATKGVGVPAVPAOOLHRENTIYEETLSKRIIPSDLSAATIKGALISLCTGCLSPSLQVNPQPKPTPIHMEAHALITLTKRVEPPVILISGCHLALVQGSDFLLKSLDIAGMDLDRKRLSLIKHRCSTMSGSADIBOLAKDGNRPHFTINPMQNAKDCSFYLAQHTDLITLHKEEBCIEYRQIISAKIDIAAIVOHATACHAKTRKTHAIFCOKNLSIPANVLTAVSGVANSIYIRVLEIYVNAATGCTLCFFPLCTDNGMIAMNGIERLRAGIAGVILHVEDIRYRKPCEPVGVDISREVAALIKVPLKVAL"			
1605. 1610			
/note="putative"			
polya_signal			
polya_site			
1622			

ORIGIN /note="putative"

## Alignment Scores:

Pred. No.: 2,38e-193 Length: 1622  
 Score: 1819.00 Matches: 351  
 Percent Similarity: 90.34% Conservative: 23  
 Best Local Similarity: 84.78% Mismatches: 40  
 Query Match: 85.60% Indels: 0  
 Gaps: 0

US-10-649-273-2 (1-414) x AK011265 (1-1622)

```

QY      1 MetLeuLeuThrLysThrAlaGlyValPhePheLysProSerLysArgHisValTyr 20
Db      207 ATGCTTAATGTTAAAGAGACAGACGAGCTATTTCCAGAGCCCCCAAGAGTAAGTTAT 266
QY      21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysL1LeuValLeu 40
Db      267 GGATTTTAAAGAAATTGAAGTTGTTTCATCCAGAACTCTCTCTTTCATTAACCTGGTCTCG 326
QY      41 GlyL1LeuThrSerCysAspAspThrAlaAlaValAspGluThrGlyAsnVal 60
Db      327 GGAATTGAAACCAAGCTGTGATGACACAGAGCGCTGTGTGATGAAACCTGGGAATGTG 386
QY      61 LeuGlyGlyAlaLeuHisSerGlnThrGlyValHisLeuLysThrGlyL1LeuValPro 80
Db      387 CTGGGGGAGACATGCTACCTCCAACTCAAGTTCACTGAAACAGTGGAGATGTTCT 446
QY      81 ProL1a1aGlnGlnLeuHisArgGlyL1a1aGlnGlnLeuValGlnGlnL1a1aLeuSer 100
Db      447 CCACTAGCTCAACAACCTTCAAGAGAAATATTCACAGAAATGTAAGAAACCTTTT 506
QY      101 AlaSerGlyValSerProSerAspLeuSerAlaAlaThrThrL1LeuProGlyLeu 120
Db      507 GCCTGTGAATCACCCCAAGCATCTTCAGCAATTCGCAATCAATCAACCGGAGACTG 566
QY      121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyL1LeuLysLys 140
Db      567 GCCCTAAGCTGGAGGTTGGCTTATCTTACCTTCAAGCTGTAATTCATGTTTAAAAAG 626
QY      141 ProPheLeuProLeuHisHisIleMetGlyL1a1aHisAlaLeuThrL1a1aArgLeuThrAsnLys 160
Db      627 CCATTCAATCCGATTCATCATGAGAGCTCAGCATGACTATTAAGGTCCCAATTA 686
QY      161 ValGluPheProPheLeuValLeuLeuL1SerGlyL1HisCysLeuLeuL1a1aLeuVal 180
Db      687 GTAGAATTTCTTTTATTGTTCTTTGATTCTGGCGGTCACTGCTGTTGGCATTTAGTC 746
QY      181 GlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspL1a1aProGlyAspMet 200
Db      747 CAAAGTGTTCCTGATTTCTCTCTGGAAGCTTGTGGACATGACACAGGCAATG 806
QY      201 LeuAspLysValAlaArgArgLeuSerLeuL1LysHisProGlyCysSerThrMetSer 220
Db      807 CTTGACAAAGGTGGCAAGAACTTTCTTATCAACATCCAAATGTTTCTCAATGAT 866
QY      221 GlyL1LysAlaLeuGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspL1LeuLys 240
Db      867 GGTGAAAAGCTATGAAACAGTTGCGCAAAAGCAAGAAATGATTCATTTTCTATCAT 926
QY      241 ProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThr 260
Db      927 CCACCTTACGAGAAAGCTTAAGAAATGTCATTTTCTTTCACAGGACTTCACAAATATTA 986
QY      261 AspLysL1LeuMetLysGlyL1a1aGlnGlyL1LeuGlyL1GlyL1a1aLeuSer 280
Db      987 GATTAAGCTTAATTAACACACAGGAAAGAAAGAGGATTTGAAGAGAGCAAAATCTGTGA 1046
QY      281 SerAlaAspL1a1aL1a1aThrValGlnHisThrMetAlaCysHisLeuValLysArg 300
Db      1047 TCACTGAGACATTTGCTGCTGGGTACAGCATGACAGGCTGCACTTCCGAAAGAA 1106
  
```

```

QY      301 ThrHisArgAlaL1LeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaVal 320
Db      1107 ACACATCGCGCTATTTCTGTTTGGCAAGAGAAAATTTCTCTCCAGCTAACGACAT 1166
QY      321 LeuValAlaSerGlyL1ValAlaSerAsnPheThrL1a1aArgAlaLeuGlnL1LeuLys 340
Db      1167 TTAGTGTATCTGAGAGGTTCTTCAAGTAACTTGATCATCCGAAACATTTGAAATGTC 1226
QY      341 ThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuLeuSerThrAspAsnGly 360
Db      1227 GCAAAATGCAACGACAGCTGTTGTGTCCACTCCCAAGACTGTGCAATGGC 1286
QY      361 IleMetL1a1aL1a1aTrpAsnGlyL1LeuArgLeuArgAlaGlyLeuGlyL1LeuHisAsp 380
Db      1287 ATCATGATTCATGGAATGGAATGGAATTAACGTCGCTGGCTGGCTTTTACATGAT 1346
QY      381 IleGlyL1LeuArgTyrGluProLysCysProLeuGlyValAspL1LeuSerGlyVal 400
Db      1347 GTAGAAGACATCCGATGATGAACCAAAATGCTCTTGGAGTGAACATATCCAGAGAA 1406
QY      401 GlyGlyAlaSerL1LeuValProGlnLeuLysMetGluL1Leu 414
Db      1407 GCAAGAGCTGCCATTAAGTACCGGATTAAGAAATGCACTT 1448

RESULT 3
BC030671 1609 bp mRNA linear HTC 19-NOV-2003
LOCUS Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA
DEFINITION Clone IMAGE:1226118), containing frame-shift errors.
ACCESSION BC030671
VERSION BC030671.1 GI:21040459
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1609)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schaller,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,L., Musunna,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abrahamson,R.D., Mulhany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Huiyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahay,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywicki,M.I., Skalska,U., Smalls,D.E.,
Scherer,B.A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL MEDLINE
PUBMED 22388257
12477932
REFERENCE 2 (bases 1 to 1609)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Marcello Bento Soares, Ph.D.
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
  
```

DNA Sequencing by: Institute for Systems Biology  
 http://www.systemsbio.org  
 contact: amadan@systemsbio.org  
 Ann Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNLT at: <http://image.llnl.gov>  
 Series: IRAX Plate: 66 Row: e Column: 10  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21312463  
 This clone has the following problem: frame shifted.

## FEATURES

Location/Qualifiers  
 1..1609  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1226118"  
 /issue\_type="Thymus gland, mouse"  
 /clone\_lib="Soares\_thymus\_2mbmt"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3-Pac"

## ORIGIN

## Alignment Scores:

Pred. No.: 1.64e-172 Length: 1609  
 Score: 1634.00 Matches: 327  
 Percent Similarity: 83.86% Conservative: 21  
 Best Local Similarity: 78.80% Mismatches: 40  
 Query Match: 76.89% Indels: 27  
 Gaps: 2

US-10-649-273-2 (1-414) x BC030671 (1-1609)

QY 1 MetLeuIleLeuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValTyr 20  
 Db 229 ATGCTAATGTTAAGAGACAGACAGACTATTCCTCCAAAGCCCAAGATTAAGTTAT 288  
 QY 21 GluPheLeuArgSerPheAsnPheHisProGlyThrLeuPheLeuHisLysIleValLeu 40  
 Db 289 GGATTTTAAAGAAATTAAGTTCATCCCAAGCTCTCTGTCATTAAGTGTCTCTG 348  
 QY 41 GlyIleGluThrSerCysAspAspThrAlaAlaValIleArgGluThrGlySerVal 60  
 Db 349 GGAATTGAACACACTGTGATGACACAGACCGCTGTGTGATGAAGAACTGGGAATGG 408  
 QY 61 LeuGlyGluAlaIleHisSerGlnThrGlyValHisLeu-LysThrGlyIleValPr 80  
 Db 409 CTGGGGAGACACGCACTCCCAACTCAGGTTCTGAAAGT-----GGGATTTGTC 462  
 QY 80 OProAlaAlaGlnGlnLeuHisArgGluAsnIleGlnArgIleValGlnGluAlaLeuSe 100  
 Db 463 TCCAGTAGCTCAACAACTTCAAGGAAATATTTCAAGATTAAGTAAAGAACTTTTC 522  
 QY 100 PAlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLe 120  
 Db 523 TGCTGTAGATCACTCCCAACGATCTCTCAGCAATTCACATCAACCAACCGGACT 582  
 QY 120 ValAlaSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuVal 140  
 Db 583 GGCCCTAAGCTTGAGAGTGGCTTATCTTACGCTTACGCTAAGTAAATCGTTTAAAA 642  
 QY 140 sProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
 Db 643 GCACATTCATCCGATTCATCAATGAGGCTCAAGCACTGACTATTAAGCTCAACCAATTA 702  
 QY 160 sValGluPheProPheLeuValIleLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 Db 703 AGTAGAATTCCTTTTAACTCTTCTTCACTTCTGCGGCTCAGTCCGTGATGCAATTA 762  
 QY 180 IgGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMe 200

Db 763 CCAAGGTTTCCGATTTCTGCTCTGGGAAGCTTTGGACATGACACAGCGCAAT 822  
 QY 200 tLeuAspLysValAlaArgArgSerLeuIleLysHisProGlnCysSerThrMetSe 220  
 Db 823 GCTTGCAAGGTTGGCAAGAACTTTCTTAACTCAACATCCGAATGTTCTCAATGAG 882  
 QY 220 rGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLys 240  
 Db 883 TGGTGAAAAGCATATGAAACAGTTGGCCAAAGCGAATATGATTCATTTACTATCA 942  
 QY 240 sProProLeuHisHisAlaLysAsnLysAspPheSerPheThrGlyLeuGlnHisValTrn 260  
 Db 943 TCCACCTATGACAGAAATGATTAAGATTCGATTTTCTTCAACGGACTTCAATATTA 1002  
 QY 260 rAspLysIleIleMetLysLysGlnLysGlnGlyIleGlnLysGlyGlnIleLeuSe 280  
 Db 1003 TGAATAGCTAATTAACACACAGGAAAAGAAAGGACATTGAAGGGGAAATCTGTG 1062  
 QY 280 rSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysAr 300  
 Db 1063 ATCAGCTGCAGCATTTGCTGCTCGGTACAGCATGCAACAGCTGCCACTTGCAGAAAG 1122  
 QY 300 gThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAlaLys 320  
 Db 1123 AACACATCGCGCTATTCTGTTTTCAGAGCAAGAAAATTTGCTCTCCAGCTAACCACT 1182  
 QY 320 lleValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArgAlaLeuGluIleLe 340  
 Db 1183 ATTAGTTGATCTGGAGGGTGTGCAAGTAACTTGATCCATCCGAAAGCATGGAAATTTG 1242  
 QY 340 uThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLysLeuCyethrAspAngi 360  
 Db 1243 CGCAATGCAAGCGACGTGACGTTGTGTCTCCTCCAGACTGAGCTGACATG 1302  
 QY 360 ylleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAs 380  
 Db 1303 CATCATGATTGCA-----TGATGTCCTCTTGAGATGATATCCAGAGAACT 1315  
 QY 380 pIleGluGlyIleArgTyrlleGluProLysCysProLeuGlyValAspIleSerLysGluVal 400  
 Db 1316 -----TGATGTCCTCTTGAGATGATATCCAGAGAACT 1350  
 QY 400 lGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 414  
 Db 1351 TGCAGAACTGTCATTAAGTACCGGATTTAAATAATGCACTT 1393

## RESULT 4

B0423651

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

B0423651 870 bp mRNA linear EST 23-MAY-2002  
 AGENCOURT\_7790948 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6065828  
 5', mRNA sequence.  
 B0423651  
 B0423651.1 GI:21118966  
 EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NIH-MGC <http://mgs.nci.nih.gov/>  
 1 (bases 1 to 870)  
 Unpublished (1999)  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: ATCC/DCPD/DTF  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNLT at: <http://image.llnl.gov>  
 Plate: LLM13342 row: 1 column: 21  
 High quality sequence stop: 710.

## FEATURES

Location/Qualifiers  
1..870  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6065828"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="N1H MGC 72"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

## ORIGIN

Alignment Scores:  
Pred. No.: 7.29e-152 Length: 870  
Score: 1447.00 Matches: 285  
Percent Similarity: 98.62% Conservative: 1  
Best Local Similarity: 98.28% Mismatches: 2  
Query Match: 68.09% Indels: 2  
DB: Gaps: 0

US-10-649-273-2 (1-414) x B0423651 (1-870)

111 AAlaIleAATrThrTlLeysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPhe 130  
1 GCAATTGCACTACCATTAACCAAGACCTTGTCTTAAGCTGGAGTGGCTTATCATTT 60  
131 SerLeuGlnLeuValGlyGlnLeuLysProPheIleProIleHisIsmetGluAla 150  
61 AGCTTACAGCTGGTGAAGAGTTAAAGACCATTCATTCATTCATCATATGAGGCT 120  
151 HisAlaLeuThrTlAargLeuThrAsnLysValGluPheProPheLeuValLeuLeu 170  
121 CATGACACTACTATTAGGTGACCAATGAAGATTTCTTTTGTAGTTCTTTGATT 180  
171 SerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGly 190  
181 TCTGGAGGTCACTGCTGTGTGATAGTTCAGAGATTCAGATTTCTGTCTTGGAA 240  
191 LysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeu 210  
241 AAGTCTTGGACATGACACAGCTGACATGCTTGAAGAGTGGCAAGAACTTCTTTA 300  
211 TlLeysHisProGlyCysSerThreSerGlyGlyLysAlaIleGlnHisLeuAlaLys 230  
301 ATAAACATCCAGAGTGTCCACCATGAGTGGGAAAGCCATGAAACATTTGGCCAAA 360  
231 GlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLysAsnCysAsp 250  
361 CAAGGAAATAGATTTTCATTTGACATCAAACTCCCTTCATCATCAATAAATGTGAT 420  
251 PheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGly 270  
421 TTTTCTTTTACCTGACCTTCAACACCTTACTGATTAATAATGAAAAGAAAAGAG 480  
271 GlnGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGln 290  
481 GAAGGATATTGAGAGAGGGGAAATCTGCTTTCAGACAGACGATGCTCCACAGTACG 540  
291 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 310  
541 CACACAAATGGCATGTCATCTTGTAAGAAACACATCGGGCTATTCTGTTGTGAAGCAG 600  
311 ArgAspLeuLeuProGlyAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsn 330  
601 AGAGACTGTTCCTCAAAATATATGACATGCTGTTGCACTCTGGTGGTCTCCAGATAC 660  
331 PheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 350  
661 TTCTATATCCGAGAGCTCTGGAATTTTAACAAAGCAACAGACATGCACTTTGGTGTCT 720

351 ProProArgLeuCysThrAspGlnGlyIleMetIleAlaTyrAsnGlyIleGluArg 370  
721 CCTCTCCAGACATATGACATGATATGACATTAATGACATGAAATGTATTAAGAAG 780  
371 LeuAlaGlnGlyLeuGlnIleLeuHisAspIleGlnGlyIleArgTyrGluProLysCys 390  
781 CTACGCTGCTGGCTTGGGCAATTTTACATGACATGAGGATCCGCTATGAACCAATGG 840  
390 sProLeuGlnValAspIleSerLys 398  
841 TCTCTCTGGAGTACATATCAAAA 866

## RESULT 5

BX391919/c 852 bp mRNA linear EST 28-APR-2004  
LOCUS BX391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
DEFINITION BX391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
CDNA clone CSODK001YB02 3-PRIME, mRNA sequence.  
ACCESSION BX391919  
VERSION BX391919.2 GI:46846154  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

## REFERENCE

AUTHORS Li W.B., Gruber C., Jesse J., and Polayes D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On May 13, 2003 this sequence version replaced gi:30611736.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1 (bases 1 to 852)  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
1240.r  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?cs=CS0BA10352B07\_CS03317\_1&c=1240.r

## FEATURES

Location/Qualifiers  
1..852  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODK001YB02"  
/cell\_type="HELA CELLS COT 25-NORMALIZED"  
/clone\_id="HELA"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:

Pred. No.: 2.41e-141 Length: 852  
Score: 1353.50 Matches: 274  
Percent Similarity: 96.14% Conservative: 0  
Best Local Similarity: 96.14% Mismatches: 8  
Query Match: 63.69% Indels: 4  
DB: Gaps: 2

US-10-649-273-2 (1-414) x BX391919 (1-852)

72 HisLeuLys---ThrGlyGlyTlLeValProProAlaAlaGlnGlnLeuHisArgGluAsn 90  
850 CATTTAACACAGGTGATGCT-----CTCCAGACGCTCAACAGCTTCAACAGAGA-AAT 798  
91 TlGlnArgIleValGlnGlnAlaLeuSerAlaSerGlyValSerProSerAspLeuSer 110

Db	797	TTTCAAGCAATAGTACCAAGAGCTCTTTCTGCAGTGNAGTCTCCCAAGTACCTCTCA	738
Qy	111	AlaIleAlaThrTThIleLysProGlyLeuAlaLeuSerLeuGlyValGlyLeuSerPhe	130
Db	737	GCAATTTGAMCTACACATAAACCCAGNACTGCTTTTAAAGCTTGAGGTGGCTTATCATTT	678
Qy	131	SerLeuGlnLeuValGlyGlnLeuLysLysProPheIleProIleHisIleMetGlnIle	150
Db	677	AGCTTACAGCTGGTGGAGACGTTTAAAGCCATTCATTTCCATTCATCATATGAGGCT	618
Qy	151	HisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuValLeuLeuIle	170
Db	617	CATGACCTTACTATTAAGTTGACCAATAAAGTAAATTTCCTTTTTAACTTTTGANT	558
Qy	171	SerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGly	190
Db	557	TCTGAGAGTCACTGCTCTGTGGCATTTAGTTCAAGAGATTTTCAGATTTCTGCTTTGGA	498
Qy	191	LysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeu	210
Db	497	AAGTTTGTGACATATGACACAGGTACATGCTTGACAGAGTGCAAGAGACTCTTTCTTA	438
Qy	211	IleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLys	230
Db	437	ATPAAACATCCAGAGTGTCTCCACCATGAGTGTGGGAAAGCATAGAACATTTGGCCAA	378
Qy	231	GlnGlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLysAsnCysAsp	250
Db	377	CAAGGAAATAGATTTTCATTTTGAACATCAACCTCCCTTGATCATGTCTAAATATGTGAT	318
Qy	251	PheSerPheThrGlnLysGlnHisValHisAspLysIleIleMetLysLysGlnLysGlu	270
Db	317	TTTTCTTTTACTGACTTCAACACGTACTGATTAATATATATATAAAAGAAAAAG	258
Qy	271	GluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGln	290
Db	257	GAAGTATTTAGAGAGGGGCAAACTCTGCTTACAGACAGACATTTGCTGCCACAGTACAG	198
Qy	291	HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln	310
Db	197	CACCAATAGGACATGTCATCTTGTAAGAAAGAACACATCGGGCTATTCGTGTTGTAGCAG	138
Qy	311	ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsn	330
Db	137	AGAGACTTGTAACCTCAAAATATGACATACGTGATCGTGGGTGGTGGCCAGGTAAAC	78
Qy	331	PheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys	350
Db	77	TTCTATATCCGACAGCTCTGGAATTTTAAACAAACGACACAGTGCATTTGTGTGT	18
Qy	351	ProProProArgLeu	355
Db	17	CCTCCTCCACAGACTA	3
RESULT 6			
BQ961028			
LOCUS	BQ961028	922 bp	mRNA linear EST 21-AUG-2002
DEFINITION	AGNC00028T 8863711 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6423902		
ACCESSION	5', mRNA sequence.		
VERSION	BQ961028		
KEYWORDS	BQ961028.1 GI:22376506		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-remail.nih.gov		
	Tissue Procurement: DCTD/DTF/gazdar		

DB	QY	Score	Percent Similarity	Best Local Similarity	Query Match	DB
US-10-649-273-2 (1-414) x BQ961028 (1-922)	80	6.75e-133	92.2	275	5	3
	3	1279.00	91.80%	Conservative: 5	16	10
	100	90.16%	Mismatches: 10	Indels: 3		
	63	60.19%	Gaps: 3			
	120					
	123					
	140					
	183					
	160					
	243					
	180					
	303					
	200					
	363					
	219					
	423					
	239					
	483					







TITLE  
G.P., Bosak, S., Rubenfield, M. and Gasbarre, L.C.  
Production of EST from cDNA libraries derived from immunologically  
activated bovine gut  
JOURNAL  
Unpublished (2004)  
COMMENT  
Contact: Tad S. Sonstegard  
Bovine Functional Genomics Laboratory  
Animal and Natural Resources Institute  
Bldg. 200 Rm2A BARC-Baet, Beltsville, MD 20705, USA  
Tel: 3015048416  
Fax: 3015048414  
Email: tads@nri.barc.ueda.gov

FEATURES  
Source  
1. .749  
Location/Qualifiers  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/strain="Holstein"  
/db\_xref="taxon:9913"  
/clone="9BOV45\_M02"  
/sex="Male"  
/tissue\_type="Pooled"  
/dev\_stage="Multiple"  
/lab\_host="DH10B T1 phage resistant"  
/clone\_1lb="BARC 9BOV"  
/note="Organ: Abomasum; Vector: phgen-1; Site 1: EcorV;  
Site 2: NotI; Equimolar amounts of mRNA extracted from  
fundic and pyloric abomasum of 18 and 21 week old steers.  
Exposure to Osterriaga osterriagi was initiated at 15 weeks  
of age. fundic and pyloric abomasum"

## ORIGIN

## Alignment Scores:

Pred. No.: 2.09e-119 Length: 749  
Score: 1158.00 Matches: 225  
Percent Similarity: 95.18% Conservative: 12  
Best Local Similarity: 90.36% Mismatches: 12  
Query Match: 54.49% Indels: 0  
DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x CK982692 (1-749)

QY 49 ThrAlaAlaValAlaValAspGluThrcGlyAsnValLeuGlyGluAlaIleHisSerGln 68  
DB 1 ACAGCAGCTGCTGATGAGTGAAGCTGGAATGTTTGGAGAGCAATACATCCCA 60  
QY 69 ThcGluValHisLeuValThcGlyGlyGlyValProProAlaAlaGlnGlnLeuHisArg 88  
DB 61 ACTGAGCTTCATTTAAACAGAGTGGGATTAATCTCCAGTAGCTCAACAGCTTACAGA 120  
QY 89 GluAsnIleGlnArgGlyLeuAlaGlnGluAlaLeuSerAlaSerGlyValSerProSerasp 108  
DB 121 GAAATATTTCAACGATAGTACAAAGAGCTCTCTGTCAGAGAGTCTCTCCAGGTAA 180  
QY 109 LeuSerAlaIleAlaThcThrIleYsProGlyLeuAlaLeuSerLeuGlyValGlyLeu 128  
DB 181 CTCACAGAGTTCACACCAACCAATAGCCAGAGCTTGTTCATTCAGCTTACAGTAA 240  
QY 129 SerPheSerLeuGlnLeuValGlyGlnLeuValYsProPheIleProIleHisHisMet 148  
DB 241 TCGTTTACCTTACCACTGAGTACCAAGTTCATTTAAAGCCCTTCATTCATTCATCA 300  
QY 149 GluAlaHisAlaLeuThrIleArgLeuThrAsnValGluPheProPheLeuValLeu 168  
DB 301 GAGGCTCATGACCTACATTAATAGGTAAACAATAGAGTAAATTCGTTTATGTTCTT 360  
QY 169 LeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 188  
DB 361 TTGATTTCTGAGGTCATTCCTTTGTCATTAAGTAAAGAGTTCAGATTTCTCTCT 420

QY 189 LeuGlyLeuSerLeuAspIleAlaProGlyAspMetLeuAspValAlaArgArgLeu 208  
DB 421 CTGGGAAGTCTTTGGACATAGCACCAGGATGATCTTCACAGTAGCAAGAACACTT 480  
QY 209 SerLeuIleYsHisProGlyCysSerThcMetSerGlyGlyValAlaIleGlnHisLeu 228  
DB 481 TCTTATTAATAACATCAGAGTCTCCACCATGATGAGCGGAGGCTATAGAACATTTTG 540  
QY 229 AlaValGlnGlyAsnArgPheHisPheAspIleLeuProProLeuHisHisAlaValAsn 248  
DB 541 GCCAAACAGGAAATGATGATTCATTTGATTCACAGCTCCCATGCAACGCTTAAAT 600  
QY 249 CysAspPheSerPheThrcGlyLeuGlnHisValThrAspValIleIleMetLysGlyLeu 268  
DB 601 TGTGATTTTCTTTTCTGACCTTCAACAGCTTATGATGATGATGATGATGATGATGAT 660  
QY 269 LysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaThr 288  
DB 661 AAAGAGGAAGTATGAGCAGGAGGCGAGGCTCTCTTCAAGCTGCAAGCATTCGTCGCG 720  
QY 289 ValGlnHisThcMetLacYsHisLeu 297  
DB 721 GTGCAGACACCGCTGCTGCCACATTT 747

RESULT 9  
LOCUS CO738006 800 bp mRNA linear EST 29-JUL-2004  
DEFINITION SLH03c19e02f1 squirrel heart library 1 Spermophilus lateralis cDNA  
clone 19e02 5', mRNA sequence.  
ACCESSION CO738006  
VERSION CO738006.1 GI:50825276  
KEYWORDS EST.  
SOURCE Spermophilus lateralis (golden-mantled ground squirrel)  
ORGANISM Spermophilus lateralis  
Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae;  
Spermophilus.

REFERENCE 1 (bases 1 to 800)  
Williams,D.R., Gracey,A.Y., Martin,S.L., Hughes,M.A., Li,W.,  
Rogers,J. and Cousins,A.R.  
Microarray analysis of transcriptional changes during hibernation  
in the golden mantled ground squirrel, Spermophilus lateralis  
Unpublished (2004)  
JOURNAL  
CONTACT: Andrew R. Cousins  
Laboratory for Environmental Gene Regulation  
University of Liverpool  
School of Biological Sciences, The Biosciences Building, Crown  
Street, Liverpool, United Kingdom, L69 7GB  
Tel: +44(0)151-795-4510  
Fax: +44(0)151-795-4431  
Email: cousins@lly.ac.uk  
Vector has been trimmed from this EST.  
Plate: 19 row: e column: 02  
Seg primer: pfic T7 (5'-AATGACCTCACTATAGG-3')  
High quality sequence stop: 800.  
Location/Qualifiers

FEATURES  
source  
1. .800  
/organism="Spermophilus lateralis"  
/mol\_type="mRNA"  
/db\_xref="taxon:76772"  
/clone="19e02"  
/sex="Male & female"  
/tissue\_type="Heart"  
/dev\_stage="Adult"  
/lab\_host="R. coli B Electromax DH10B"  
/clone\_1lb="squirrel heart library 1"  
/note="Vector: pFUC; Site 1: SalI GCGAG; Site 2: BamHI  
GATTC; NotI; linearized and subcloned cDNA library prepared  
from heart of hibernating and summer animals"

## ORIGIN

Alignment Scores:

Pred. No.: 8.74e-118 Length: 800  
 Score: 1144.00 Matches: 226  
 Percent Similarity: 97.05% Conservative: 4  
 Best Local Similarity: 95.36% Mismatches: 7  
 Query Match: 53.84% Indels: 0  
 DB: 7 Gaps: 0

US-10-649-273-2 (1-414) x C0738006 (1-800)

QY 1 MetLeuileLeuThrLyThrAlaGlyValPhePheLyProSerLyAspGlyValTyr 20  
 Db 88 ATGCTAATTTGAATTAAGACACCGGAGTGTCTTTTAAACCATTCGAAAGAAAGTTGAT 147  
 QY 21 GluPheLeuAspSerPheAsnPheHisProGlyThrLeuPheLeuHisValLeu 40  
 Db 148 GGATTTTAAAGAGTTTAAATTTTCACTCGAACAATATTTCTCATTAATAATGATTG 207  
 QY 41 GlyIleGlyThrSerCysAspAspThrAlaAlaAlaValAspGlyThrGlyValAsnVal 60  
 Db 208 GGTATTGAAACAGCTGTGATGATACAGACGCTGGGTGTGATGAATGAATGGAATGTG 267  
 QY 61 LeuGlyGlyAlaIleHisSerGlnThrGlyValHisLeuLyThrGlyGlyIleValPro 80  
 Db 268 TTGGGAGAGCAATACATTTCCCAACTGAAAGTTCATTTAAAAACAGTGGGATTTATTCCT 327  
 QY 81 ProAlaIleGlnGlnLeuHisArgGlyAsnIleGlnArgIleValGlnGlyAlaLeuSer 100  
 Db 328 CCAGTAGCTCAACATTCATTAAGAAACATTCACGAAATGACAAAGAGCTGTTCCT 387  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLySerProGlyLeu 120  
 Db 388 GCCAGTGAATATCTCCAAAGTGAATCTCAGCGATTCGACATCACTAATAAAGCAGAGCTT 447  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLySyls 140  
 Db 448 GCTTTAAGCTTAGAGGTAGGCTTATCATTTAGTTTACAGCTGTGTGACACGCTGAAAAAG 507  
 QY 141 ProPheIleProIleHisHisMetGlyAlaHisAlaLeuThrIleArgLeuThrAsnLySyls 160  
 Db 508 CCATTTCATCCCATTCATCATTAAGAGGCTCAGCACTTACTATTAGGTGACCAATATAA 567  
 QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
 Db 568 GTAGAGTTTCCATTTTATGTTCTTTGATTTCTGAGGTCAGTCTCTGTGCAATTAGTT 627  
 QY 181 GlnGlyValSerAspPheLeuLeuGlnGlyLySerLeuAspIleAlaProGlyAspPheT 200  
 Db 628 CAAGGAGTTTCAGATTTTCTGCTCTTGGAAGTCTTTGACATTCGACACGATGACATG 687  
 QY 201 LeuAspLyValAlaArgArgLeuSerLeuIleLySerHisProGlyCysSerThrMetSer 220  
 Db 688 CTTTACCAAGGTAGCAAGAAAGACTTTCTTTAAACATTCGAAATGCTCCACATGAGT 747  
 QY 221 GlyGlyLyValAlaIleGlnHisLeuAlaLySerGlnGlyAsnArgPheHisPhe 237  
 Db 748 GCGCGGAAGGCTTATGAAACATTTGGCCAAACAGAAACAGATTTTCATTTT 798  
 RESULT 10  
 LOCUS BM907988 1109 bp mRNA linear EST 12-MAR-2002  
 DEFINITION AGENCOURT\_6707465 NIH\_MGC\_119 Homo sapiens cDNA clone IMAGE:5744934  
 ACCESSION BM907988  
 VERSION BM907988.1 GI:19358367  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1109)  
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed By: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing By: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLM12767 row: b column: 07  
 High quality sequence stop: 602.  
 Location/Qualifiers  
 1..1109

## FEATURES

## source

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5744934"  
 /tissue\_type="medulla"  
 /lab\_host="DH10B"  
 /clone\_idb="NIH\_MGC\_119"  
 /note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: EcoRV (destroyed); RNA source normal medulla from  
 anonymous male age 27. Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.3 kb, insert size range  
 0.9-3 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research genetics tracking code 013. Note:  
 this is a NIH\_MGC Library."

## ORIGIN

## Alignment Scores:

Pred. No.: 3.28e-116 Length: 1109  
 Score: 1132.00 Matches: 241  
 Percent Similarity: 88.89% Conservative: 15  
 Best Local Similarity: 83.68% Mismatches: 23  
 Query Match: 53.27% Indels: 9  
 DB: 5 Gaps: 4

US-10-649-273-2 (1-414) x BM907988 (1-1109)

QY 1 MetLeuileLeuThrLyThrAlaGlyValPhePheLyProSerLyAspGlyValTyr 20  
 Db 129 ATGCTAATTTGAATTAAGACACCGGAGTGTCTTTTAAACCATTCGAAAGAAAGTTAT 188  
 QY 21 GluPheLeuAspSerPheAsnPheHisProGlyThrLeuPheLeuHisValLeu 40  
 Db 189 GAATTTTAAAGAGTTTAAATTTTCACTCGAACACTAATTTCTCATTAATAATGATTTG 248  
 QY 41 GlyIleGlyThrSerCysAspAspThrAlaAlaAlaValAspGlyThrGlyValAsnVal 60  
 Db 249 GGAAATTTGAAACTAGTGTATGATATACAGAGCTGTGTGTGATGAATGGAATGAAATGTG 308  
 QY 61 LeuGlyGlyAlaIleHisSerGlnThrGlyValHisLeuLyThrGlyGlyIleValPro 80  
 Db 309 TTGGGAGAGCAATACATTTCCAAACTGAAGTTCAATTTAAAAACAGTGGGATTTGTTCT 368  
 QY 81 ProAlaIleGlnGlnLeuHisArgGlyAsnIleGlnArgIleValGlnGlyAlaLeuSer 100  
 Db 369 CCAGGAGCTCAACAGCTTCACAGAGAAATATTCAACGAATAGTACAGAAAGCTTTTCT 428  
 QY 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLySerProGlyLeu 120  
 Db 429 GCCAGTGGAGTCTTCCAAAGTACCTCTCAGCAATTCGCAATTCATTAACCAAGAGCTT 488  
 QY 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLySyls 140  
 Db 489 GCTTTAAGCTTAGAGGTAGGCTTATCATTTAGCTTACAGCTGTGTGACAGCTTAAAAAG 548  
 QY 141 ProPheIleProIleHisHisMetGlyAlaHisAlaLeuThrIleArgLeuThrAsnLySyls 160  
 Db 549 CCATTTCATCCCATTCATCATTAAGAGGCTCAGCACTTACTATTAGTTGACCAATATAA 608

QY 161 ValGluPheProPheLeuValLeuLeuIleSerGlyIleCysLeuLeuAlaLeuVal 180  
 Db 609 GTAGAAATTCCTTTTATGTTCTTTGATTTCTGGAGGTCACTGCTGTGGCATAGTT 668  
 QY 181 GlnGlyValSerAspPheLeuLeuLeuGlyIleSerLeuAspIleAlaProGlyAspMet 200  
 Db 669 CAAGAGATTCATGATTTCTGCTCTGAAAGGCTTGGACATACACACAGGGGACATG 728  
 QY 201 LeuAspLeuValAlaArgArgLeuSerLeuIleLeuHisProGlyCysSerThrMetSer 220  
 Db 729 CTGTAACACAGGGGACAGAAAGATTTCTTTATTAACATCCAGAGGCTCCACCATAGG 788  
 QY 221 GlyIleValAlaIleGluHisLeu--AlaIleGlnGlyAsnArgPheHisPhe--Asp 238  
 Db 789 GGGGGGAAAGACCATTAACCATTTGGGCAACCAAGAAATAGATTTCTTTGGAAAT 848  
 QY 239 IleIleProPheLeuHisHis-AlaIleAsnCysAspPheSerPheThrGly--LeuGln 257  
 Db 849 CCAAACTCCCTTCATATGTTAAATAATGGGATTTTCTTTAATGGGAATTTCAA 908  
 QY 258 HisVal--ThrAspLeuIleIleMetIleIleGlyIleGlyIleGlu----- 274  
 Db 909 CACGCTTACTGCTTAATAATATGAAAGGAAAGGAAAGGAAAGGAAATTTGGGA 968  
 QY 275 --LysGlyGlnIleLeu 279  
 Db 969 AAAGGGGCAAAATCTCTG 986

RESULT 11  
 LOCUS CF114247 765 bp mRNA linear EST 23-JUL-2003  
 DEFINITION Shlutfomica07498 Rat lung airway and parenchyma cDNA libraries  
 VERSION CF114247  
 KEYWORDS Rattus norvegicus cDNA clone NA4935 5', mRNA sequence.  
 SOURCE CF114247.1 GI:33173962  
 ORGANISM Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus

REFERENCE 1 (bases 1 to 765)  
 AUTHORS Shultz,M.A., Zhang,L., Gu,Y.-Z., Baker,G.L., Fannuchi,M.V.,  
 Padua,A.M., Gurske,W.A., Morin,D., Penn,S.G., Jovanovich,S.B.,  
 Plopper,C.G. and Buckpitt,A.R.  
 TITLE Gene expression analysis in response to lung toxicants: I.  
 JOURNAL Sequencing and microarray development  
 COMMENT Am. J. Respir. Cell Mol. Biol. 30 (3), 296-310 (2004)  
 CONTACT: Shultz MA  
 DEPT. of Molecular Biosciences, School of Veterinary Medicine  
 University of California, Davis  
 111 Haring Hall, One Shields Avenue, Davis, CA 95616, USA  
 TEL: 530 752 0793  
 FAX: 530 752 4698  
 EMAIL: mshultz@ucdavis.edu

Average Phred score is 20 or better. All poor quality data (Phred <  
 20) and vector/linker sequence has been removed.  
 High quality sequence stop: 765.

FEATURES  
 source 1..765  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="NA4935"  
 /sex="male"  
 /tissue\_type="airway or parenchyma"  
 /dev\_stage="adult"  
 /clone\_id="Rat lung airway and parenchyma cDNA libraries"  
 /note="Organ: lung; Vector: pGEM-Ti(+) Site 1: Eco RI;  
 Site 2: Not I; mRNA was isolated from microdissected rat  
 lung airways and parenchyma tissues."

ORIGIN

Alignment Scores:  
 Prod. No.: 4 26e-114 Length: 765  
 Score: 1111.00 Matches: 216  
 Percent Similarity: 93.78% Conservative: 10  
 Best Local Similarity: 89.63% Mismatches: 15  
 Query Match: 52.28% Gaps: 0  
 DB: 7

US-10-649-273-2 (1-414) x CF114247 (1-765)

QY 24 ArgSerPheAsnPheHisProGlyThrLeuPheLeuHisIleValIleGlu 43  
 Db 39 AAGAAATTAATGTTCACCTAGACGACCTTTTCATCATTAACGTGGGAAATTGAA 98  
 QY 44 ThrSerCysAspPheThrAlaAlaValIleAspGluThrGlyAsnValIleGlyGlu 63  
 Db 99 ACCAGCTGTGATACACAGACGCTGTGTGGATGAAGAAATGTGCTGGAGAA 158  
 QY 64 AlaIleHisSerGlnThrGluValHisLeuIleThrGlyIleValIleProProAla 83  
 Db 159 GCACTCACTCCCAAGAGAGAGTCAATCTGAAGACAGTGGAGTTCTTCCAGTCGCT 218  
 QY 84 GlnGlnLeuHisArgGluAsnIleGlnArgIleValIleGlnIleValIleSerAlaSer 103  
 Db 219 CAACAGCTTCACAGAGAAATATTCAGCGAATGTAGAGAAAGCTCTTCTGCCAGTGG 278  
 QY 104 ValSerProSerAspLeuSerAlaIleAlaThrThrIleIleProGlyLeuAlaSer 123  
 Db 279 GTCTCCCAAGGACCTCTCAGCAATTTGCAATCATCAACACAGAGCTGGCCCTAAGC 338  
 QY 124 LeuGlyValGlyLeuSerPheSerLeuGlnIleValGlyGlnLeuIleValIleProPhe 143  
 Db 339 CTGGGGGTGGCTTATCTTTAGCGTACAGTAAATCAATTTAAAAAGCAATTTATC 398  
 QY 144 ProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeuValIle 163  
 Db 399 CCATATCATCATAGAGGCTCAGCGCTGACATTTAGGCTACCCACAAAGTCGATTT 458  
 QY 164 ProPheLeuValLeuLeuIleSerGlyIleHisCysLeuLeuAlaLeuValGlnIleVal 183  
 Db 459 CTTTATTTAGTTCTTTGATTTCTGGAGGCCACTGCTGTGGCGTTAGTTCAGAGTGT 518  
 QY 184 SerAspPheLeuLeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIle 203  
 Db 519 TCAGATTTTCTCTCTCTCGGAAAGTCCCTGACATGACCCAGGCCACATGCTTGA 578  
 QY 204 ValAlaArgArgLeuSerLeuIleIleHisProGlyCysSerThrMetSerGlyIleVal 223  
 Db 579 GTGGCAAGAGACTTTCTTTATTCAAACATCCAGAAATGTTCTTACATAGAGTGGGAAA 638  
 QY 224 AlaIleGlnHisLeuAlaIleGlnGlyAsnArgPheHisPheAspIleIleSerProLeu 243  
 Db 639 GCTATAGAAATTTGGCCAAAGAGAAATAGATTTCACTTACTATCATCATCACCAG 698  
 QY 244 HisHisAlaIleAsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspIle 263  
 Db 699 CAGAATGTCAAGAACTGTATTTTCTTTAGCGGACTTCAACATGTCAACCGATAGCTA 758  
 QY 264 Ile 264  
 Db 759 ATA 761

RESULT 12  
 LOCUS B0636028 640 bp mRNA linear EST 15-JUL-2002  
 DEFINITION h03d11.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he  
 B0636028  
 ACCESSION B0636028  
 VERSION B0636028.1 GI:21760487  
 KEYWORDS EST  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
1 (bases 1 to 640)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS  
Wistow, G., Bernstein, S.L., Wyatt, M.K., Ray, S., Behal, A., Touchman, J.W., Bouffard, G., Smith, D., and Peterson, K. Expressed sequence tag analysis of human retina for the NEIRank project: Retbindin, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts  
JOURNAL  
MEDLINE  
PMID  
12107411  
Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 03 row: d column: 11  
Seq primer: M13RPI reverse primer (ABI).  
FEATURES  
Source  
1..640  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="hd03d11"  
/tissue\_type="Retina"  
/dev\_stage="Adult"  
/lab\_host="EMD10B"  
/clone\_lib="Human Retina cDNA (un-normalized, unamplified): hd/he"  
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor (5'-TGACTGAGTTCAGATCGGAGCGGCCG(T)15-3'). EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN  
Alignment Scores:  
Pred. No.: 1,228-112 Length: 640  
Score: 1097.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 51.62% Indels: 0  
DB: 5 Gaps: 0  
US-10-649-273-2 (1-414) x B0636028 (1-640)

QY 206 ArgArgLeuSerLeuIleIyShiSPROGUySserThMetSerGIyLySalaIle 225  
Db 3 AGAAGACTTTCTTTATATAAATCATTCACAGTCCTCACCATGAGTGTGGAAAGCCATA 62  
QY 226 GlnHisLeuAlaIyGlnGlyVasnaRgPhHisSphEasPILeIySPROProLeuHISHis 245  
Db 63 GAACATTTGGCCAAACAGGAATAGATTTCATTTGACATCAAACTCCCTTGCAATCAT 122  
QY 246 AlaIySaAnCyAspPhSerPheThrgIyLeuGlnHisValThraSphIyIleIleMet 265  
Db 123 GCTAAAAATTGATTTTCTTTACTGACCTTCAACAGCTTACTGATTAATAATG 182  
QY 266 LysIySGIuYSGIuGlnGlyIleGIuYSGIyGlnIleLeuSerSerAlaIaapIle 285  
Db 183 AAAAAAGAAAAAGAGAGTATTGAGAAAGGGCAAAATCCTGCTTCAACAGACATTT 242

QY 286 AlaAlaThrValGlnHisThrMetAlaCySHISLeuValIySArgThrHisArgAlaIle 305  
Db 243 GGTGCACAGTACAGACACAAATGGCATCTTGTGAAAAGAACACATGGGCTATT 302  
QY 306 LeupHeCyELySGIuNArGAspLeuLeuPProGlnaenAenAlaValLeuValAlaSerGly 325  
Db 303 CGTTTTGTAGACAGAGACATCTTCACTCAAAAATAGCAGTACGTGTCATCTGGT 362  
QY 326 GIVAlAlaSerAspHeThyIleArgArgAlaLeuGlnIleLeuThraSnaIaThrgIn 345  
Db 363 GGATGCAGAGTACTTCTATATCCGAGAGCTTGAAATTTTAAACAAGCAACAG 422  
QY 346 CyThrLeuLeuCySPROProPArgLeuCyThraSAsnGlyIleMetIleAlaTrp 365  
Db 423 TCACATTTTGTGTCTCTCTCCACATATGACATATGACATTAATGATTCATGG 482  
QY 366 AaenGlyIleGIuArGLeuArGAlaGlyLeuGlyIleLeuHisAspIleGIuGlyIleArg 385  
Db 483 AATGTATTGAAAGATACGTGCTGGCATTTTACATGACATAGAGGCATCCGC 542  
QY 386 TyrGluProIyCySPROLeuGIyValaSPIleSerIySGIuValGlyIleAlaSerIle 405  
Db 543 TATGAACCAAAATGTCCTCTTGAGTACATATCAAAAGAGTTGAGAGCTTCATA 602  
QY 406 LysValProGlnLeuIySmetGluIle 414  
Db 603 AAAATACCAATTAATAATGAGATA 629

RESULT 13  
CN823245 730 bp mRNA linear EST 02-JUN-2004  
LOCUS  
DEFINITION  
Oa splbn\_04N08 M13reverse Sheep spleen\brain pSPORT1 library Ovis aries cDNA clone Oa\_splbn\_04N08 5', mRNA sequence.  
ACCESSION  
CN823245  
VERSION  
CN823245.1 GI:47951314  
KEYWORDS  
EST.  
SOURCE  
Ovis aries (sheep)  
ORGANISM  
Ovis aries  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Caprinae; Ovis.  
1 (bases 1 to 730)  
REFERENCE  
AUTHORS  
Gossner, A. and Hopkins, J.  
TITLE  
Ovine spleen/brain cDNA library  
JOURNAL  
Unpublished (2004)  
COMMENT  
Contact: J Hopkins  
Veterinary Biomedical Sciences  
University of Edinburgh  
Summerhall Square, Edinburgh, EH9 1QH.  
Email: j.hopkins@ed.ac.uk  
Plate: 04 row: N column: 08  
Seq primer: M13reverse  
High quality sequence start: 6  
High quality sequence stop: 550.  
FEATURES  
source  
1..730  
/organism="Ovis aries"  
/mol\_type="mRNA"  
/db\_xref="taxon:9940"  
/clone="Oa splbn\_04N08"  
/clone\_lib="Sheep spleen\brain pSPORT1 library"  
/note="Vector: pSPORT1"

ORIGIN  
Alignment Scores:  
Pred. No.: 1,086-106 Length: 730  
Score: 1045.00 Matches: 204  
Percent Similarity: 93.51% Conservative: 12  
Best Local Similarity: 88.31% Mismatches: 15  
Query Match: 49.18% Indels: 0  
DB: 7 Gaps: 0  
US-10-649-273-2 (1-414) x CN823245 (1-730)

```

QY 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
DB 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

```

```

FEATURES
  source
    1. .661
    /organism="Bos taurus"
    /mol_type="mRNA"
    /strain="Holstein"
    /db_xref="taxon:9913"
    /clone="8BOV_20020"
    /sex="Female"
    /tissue_type="Epithelial, Muscle"
    /dev_stage="lactating, Neonatal"
    /lab_host="DH10B Tona"
    /clone_11b="BARC 8BOV"
    /note="Organ: Intestine; Vector: pCMVSPORT6.1, Site_1:
    NotI; Site_2: EcoRI; Normalized cow cDNA intestinal
    library in pCMVSPORT6.1, constructed from equimolar mRNA
    pools derived from 5 sources, 4 lactating intestinal, 1
    neonatal intestinal 4/5 lactating, Proximal Duodenum,
    Jejunum, Distal Ileum, Colon, 1/5 Neonatal, Proximal
    Duodenum, Jejunum, Distal Ileum"

ORIGIN
Alignment Scores:
Pred. No.: 4,98e-106 Length: 661
Score: 1038.50 Matches: 203
Percent Similarity: 95.48% Conservative: 8
Best Local Similarity: 91.86% Mismatches: 9
Query Match: 48.87% Indels: 1
DB: 7 Gaps: 1

US-10-649-273-2 (1-414) x CK83139 (1-661)
QY 30 ProGlyThrLeuPheLeuHisIleValIleuGlyIleGluThrSerCyAspAspThr 49
DB 2 CCGGG--ATATTCCTTATTAACAGTATTGGGAATGAAACAGTGTGATGATACA 58
QY 50 AlaIleValIleValIleValIleValIleValIleValIleValIleValIleVal 69
DB 59 GCAAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 118
QY 70 GluValIleLeuIleValIleValIleValIleValIleValIleValIleValIle 89
DB 119 GAAGTTCATTAAACAGGTGGGATTAATTCCTCCAGTACCTCAACAGCTTCATAGGAA 178
QY 90 AAmIleGlnArgIleValIleGlnIleValIleSerIleValIleSerProSerAspLeu 109
DB 179 AATATTCAACGCACTAGTACAGAGCTCTCTGCGCAGTCAAGTCTTCCAAAGTGAATC 238
QY 110 SerAlaIleAlaThrThrIleLeuProGlyLeuAlaLeuSerLeuGlyValIleLeuSer 129
DB 239 TCAGCAGTTGCAACCACTAAAGCCAGAGACTGTTAAAGCTTGGCGTAGGTTATCG 298
QY 130 PheSerLeuGlnLeuValIleGlnIleValIleValIleValIleValIleValIle 149
DB 299 TTTAGCTTCAACCTGTAGACCAAGTTTAAACCCCTTCAATTCATTCACATGAGAG 358
QY 150 AlaHisAlaLeuThrIleArgLeuThrAsnIleValIleGluPheProPheLeuValIleLeu 169
DB 359 GCTCATGCACTCACTTATGAGTTAAACAAATAGGTAATTCGTTTATGATCTTTTG 418
QY 170 IleSerGlyIleHisCyLeuLeuAlaLeuValIleGlnIleValIleSerAspPheLeuLeu 189
DB 419 ATTTCGAGAGCTCATGCTTTTGGCATTAAGTTAGAGAGGATTCAGATTTCTCTTCTT 478
QY 190 GlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValIleArgIleLeuSer 209

```

Db 479 GGGAATCTTTGACATAGACACAGGTGACATGCTTGACAAAGTAGCAAGACTTTC 538  
Qy 210 LeuileuHSHSPROGUyCSeThrMetSerGlyValAlaIleGluHisIleuVal 229  
Db 539 TTAATTAATAACATCCAGAGTGTCTCCACCAATGAGTGGGAGGCTATAGAACATTTGGCC 598  
Qy 230 LysGlnGlyAaenArGpHeHisPheAspIleuSPROLeuHSHISAlaIysAaenCys 249  
Db 599 AAACAGGAATATAGATTGCATTTCATTTCACAGCTCCCATGCAACGTCTAAATATGT 658  
Qy 250 Asp 250  
Db 659 GAT 661  
RESULT 15  
BQ433135 879 bp mRNA linear EST 24-MAY-2002  
LOCUS AGENCOURT 7760756 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:6016098  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ433135  
VERSION BQ433135.1 GI:21172211  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 879)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: rgsb@nsl.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LNLMI3213 row: d column: 19  
High quality sequence stop: 674.  
FEATURES  
source location/Qualifiers  
1..879  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6016098"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="RDH108 (phage-resistant)"  
/clone\_lib="NIH\_MGC\_92"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.39e-103 Length: 879  
Score: 1018.50 Matches: 210  
Percent Similarity: 94.67% Conservative: 3  
Best Local Similarity: 93.33% Mismatches: 9  
Query Match: 47.93% Indels: 3  
DB: 5 Gaps: 1  
US-10-649-273-2 (1-414) x BQ433135 (1-879)  
Qy 1 MetLeuileuThrLysThrAlaGlyValPhePheLysProSerLysArgLysValIyr 20  
Db 201 ATGCTAAATCTGACTAAGACTGACAGAGTCTTTTAAACCAATCAAAAGAAAGTTTAT 260  
Qy 21 GluPheLeuArgSerPheAsnPheHisPheProGlyThrIleuPheLeuHisLysAlleuVal 40

Db 261 GAATTTTAAAGATTTTTAATTTCATCTCGAACAATATTTCTTCATAAATAATGATTTC 320  
Qy 41 GYIIEgluThrSerCyseAspThrAlaIleValIleAspGluThrGlyAaenVal 60  
Db 321 GGAATTTGAAATAGATGTGTATATATACAGACGCTGCTGTGTGATGATAAATGAG 380  
Qy 61 LeuGlyGluAlaIleHisSerGlnThrGluValHisLysLysThrGlyIleValPro 80  
Db 381 TTGGAGAAAGCAATATCATTTCCAAACTGAAGTTCATTTAAACACAGTGGATTTGTTCT 440  
Qy 81 ProAlaIleGlnGlnLeuHisArgGluAaenIleGlnArgIleValGlnGluAlaLeuSer 100  
Db 441 CCAGCAGCTCAACAGCTTCACAGAAATATTCACGAATAGTACAGAAAGCTCTTCT 500  
Qy 101 AlaSerGlyValSerProSerAspLeuSerAlaIleAlaThrThrIleLysProGlyLeu 120  
Db 501 GCCAGAGAGTCTCTCCAAAGTACCTCTGACGAATTCACATACATTAACAGGACTT 560  
Qy 121 AlaLeuSerLeuGlyValGlyLeuSerPheSerLeuGlnLeuValGlyGlnLeuLys 140  
Db 561 GCTTTAAGCTCGGAGTGGGCTTATCATTTACCTTACCTGATGAGACAGTTAAAAAAG 620  
Qy 141 ProPheIleProIleHisIleMetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys 160  
Db 621 CATTCATTCATCCATTCATCATATGAGGCTCATGCACTTACTATTAAGTTGACCAATAAA 680  
Qy 161 ValGluPheProPheLeuValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuVal 180  
Db 681 GTAGAAATTCCTTTTAACTTCTTTGATTTCTGGAGGTCACATGTCGTGGGCAATAGTT 740  
Qy 181 GlnGlyValSerAspPheLeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMet 200  
Db 741 CAAAGATTTCAGATTTTCTGCTCTTGGAAGATCTTTGACATACACAGATGACATG 800  
Qy 201 -LeuAspLysValAlaArgArgLeuSerLeuLeu--LysHis-ProGluCysSerThrm 219  
Db 801 GCTTGCAAGGTGGGCAAGAAAGACTTCTTTNATNANAATTCACAAAGTCTCCCCC 860  
Qy 219 etSerGlyGly 222  
Db 861 ATGGAAGGGT 871

Search completed: February 16, 2005, 21:04:35  
Job time : 4943.95 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 10:46:17 ; Search time 70.2 Seconds  
(without alignments)  
3019.954 Million cell updates/sec

Title: US-10-649-273-2

Perfect score: 2125

Sequence: 1 MLILTKTAGVFPKSKRKRY.....DISKEVGASIKVPLQKMEI 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: UniProt\_03:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2125	100.0	414	2	096EV9
2	2090.5	98.4	439	2	09H4B0
3	1845	86.8	364	2	096NH5
4	1835	86.4	414	2	096EBA
5	1827	86.0	414	2	09BLB6
6	1819	85.6	414	2	09D0N0
7	1725	81.2	467	2	06AYN7
8	1283	60.4	404	2	08JFW3
9	1279	60.2	404	2	08JFR7
10	772.5	36.4	401	2	07Q9I8
11	714.5	33.6	409	2	09VWD6
12	668.5	31.5	480	2	022145
13	577.5	27.2	335	2	073H71
14	568.5	26.8	335	2	092LH8
15	564	26.5	362	2	098E16
16	558	26.3	387	2	098E16
17	553	26.0	387	1	GCP_RICPR
18	548	25.8	365	2	08UC47
19	548	25.8	366	2	07CMJ8
20	541.5	25.5	344	2	07PAG7
21	540.5	25.4	344	2	092JK6
22	539.5	25.4	359	2	08RYJ1
23	539.5	25.4	359	2	08RYJ1
24	538.5	25.3	323	2	0960S6
25	535.5	25.2	367	2	09ABZ9
26	531.5	25.0	389	2	09MD54
27	524	24.7	421	2	09J170
28	520	24.5	340	2	09FCK9
29	516.5	24.3	335	1	GCP_AQUAB
30	515.5	24.3	346	2	07VXN4
31	512	24.1	364	2	06GIR3

32	510.5	24.0	346	2	07M668	07M668 bartonella
33	510	24.0	339	2	06LV10	06LV10 photobacter
34	509.5	24.0	346	2	07W134	07W134 bartonella
35	507.5	23.9	357	2	089WV1	089WV1 bradyrhizob
36	503.5	23.7	341	2	07NUE3	07NUE3 chromobacter
37	494.5	23.3	343	2	09CLJ1	09CLJ1 pasteurella
38	494	23.2	341	2	08ESI6	08ESI6 oceanobacil
39	492.5	23.2	337	1	GCP_SALTY	P40731 salmonella
40	492.5	23.2	341	2	0915V7	0915V7 pseudomonas
41	491	23.1	364	2	06FPY1	06FPY1 bartonella
42	490.5	23.1	357	2	08FDC6	08FDC6 escherichia
43	489	23.0	353	2	07VQO9	07VQO9 candidatus
44	488.5	23.0	337	2	08XBK3	08XBK3 escherichia
45	486.5	22.9	337	2	082XN2	082XN2 nitrosomona

## ALIGNMENTS

RESULT 1	ID	096EV9	PRELIMINARY;	PRT;	414 AA.
AC	096EV9				
DT	01-DEC-2001 (Tremblrel. 19, Created)				
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)				
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)				
DE	OSGEP1 protein.				
OS	Homo sapiens (human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISUS=Muscle;				
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;				
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,				
RA	Stapleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Bromstein M.J., Ueda T.B., Tohiyuki S., Carninci P., Prange C.,				
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gutarane P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.V., Huiyk S.W.,				
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,				
RA	Krzyvinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,				
RA	Jones S.J., Marra M.A.,				
RT	Generation and initial analysis of more than 15,000 full-length human				
RT	and mouse cDNA sequences."				
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISUS=Muscle;				
RA	Strausberg R.,				
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.				
EMBL	BC011904; AAH11904.1; -.				
DR	MEKOPS; M22.004; -.				
DR	GO; GO:0008450; P:O-staolglycoprotein endopeptidase activity; IEA.				
DR	GO; GO:0008270; P:zinc ion binding; IEA.				
DR	GO; GO:0005508; P:proteolysis and peptidolysis; IEA.				
DR	InterPro; IPR000905; Peptidase_M22				
DR	InterPro; IPR009180; Pept_M22_Ostia1g1.				
DR	Pfam; PF00814; Peptidase_M22; 1.				
DR	Pfam; Pfam04537; Ostia1g1_Pf04537; 1.				
DR	PRINTS; PR00789; OSIALOPASR.				
DR	ProDom; PD002367; Peptidase_M22; 1.				
DR	TIGRfam; TIGR00329; gcp_1.				
DR	SEQUENCE 414 AA; 45122 MW; A536B333F5C6B8DD CRC64;				

Query Match 100.0%; Score 2125; DB 2; Length 414;  
Best Local Similarity 100.0%; Pred. No. 2.6e-155;  
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPKSKRYEFLRSFNPHGTLFLHKIVLGIGTSCDDTAAAVVDETGVN 60  
DB 1 MLILTKTAGVFPKSKRYEFLRSFNPHGTLFLHKIVLGIGTSCDDTAAAVVDETGVN 60  
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIQVEALSASGVSPSDISAITTIKPGI 120  
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIQVEALSASGVSPSDISAITTIKPGI 120  
QY 121 ALSIGVGSFSLQVLGOLKKPFIPIHMEAHALTIRLTNKEVPPFVLLISGHCILALV 180  
DB 121 ALSIGVGSFSLQVLGOLKKPFIPIHMEAHALTIRLTNKEVPPFVLLISGHCILALV 180  
QY 181 QGVSDPFLIGKSLDIPAGMDLKVARRLSLKHPECSSTWSGKAIETHLAKQGRFPHDIK 240  
DB 181 QGVSDPFLIGKSLDIPAGMDLKVARRLSLKHPECSSTWSGKAIETHLAKQGRFPHDIK 240  
QY 241 PPLHAKNCDFSPFTGLQHTYDKIIMKKEKEBEGIEKGQILSSAADIATVQHTMACILYR 300  
DB 241 PPLHAKNCDFSPFTGLQHTYDKIIMKKEKEBEGIEKGQILSSAADIATVQHTMACILYR 300  
QY 301 THRAILPCKORDLPPONNAVLVASGVASNFYIRRALETLTNAQCTLLCPPRLCTDNG 360  
DB 301 THRAILPCKORDLPPONNAVLVASGVASNFYIRRALETLTNAQCTLLCPPRLCTDNG 360  
QY 361 IMIWMNGIERLRAGLGIHDIIEGIRYEPKPLGVDISKEVGEASIKVPOLKMEI 414  
DB 361 IMIWMNGIERLRAGLGIHDIIEGIRYEPKPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 2  
Q9H4B0 PRELIMINARY; PRT; 439 AA.  
AC Q9H4B0, -  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Putative sialoglycoprotein type 2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=uterus;  
RA Chen J.M., Fortunato M., Barrett A.J.;  
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.  
DR MEROPS; M2.004; -  
DR Genew; HGNC:23075; OSGBPL.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008223; F:peptide activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000905; Peptidase M22.  
DR InterPro; IPR009180; Pept M22 Osi1gl.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PIRSF; PIRSF004537; Osi1glc\_ptide; 1.  
DR PRINTS; PR00789; Osi1glc\_ptide; 1.  
DR PRODOM; PD002367; Peptidase M22; 1.  
DR TIGRPFAMs; TIGR00329; gcp; 1.  
KM Protease  
SQ SEQUENCE 439 AA; 48040 MW; 44849372C784B41F CRC64;

Query Match 98.4%; Score 2090.5; DB 2; Length 439;  
Best Local Similarity 93.8%; Pred. No. 1.3e-152;  
Matches 412; Conservative 0; Mismatches 2; Indels 25; Gaps 1;

QY 1 MLILTKTAGVFPKSKRYEFLRSFNPHGTLFLHKIVLGIGTSCDDTAAAVVDETGVN 60

DB 1 MLILTKTAGVFPKSKRYEFLRSFNPHGTLFLHKIVLGIGTSCDDTAAAVVDETGVN 60  
QY 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIQVEALSASGVSPSDISAITTIKPGI 120  
DB 61 LGEAHSQTEVHLKTGGIVPPAAQOLHRENIQRIQVEALSASGVSPSDISAITTIKPGI 120  
QY 121 ALSIGVGSFSLQVLGOLKKPFIPIHMEAHALTIRLTNKEVPPFVLLISGHCILALV 180  
DB 121 ALSIGVGSFSLQVLGOLKKPFIPIHMEAHALTIRLTNKEVPPFVLLISGHCILALV 180  
QY 181 QGVSDPFLIGKSLDIPAGMDLKVARRLSLKHPECSSTWSGKAIETHLAKQGRFPHDIK 240  
DB 181 QGVSDPFLIGKSLDIPAGMDLKVARRLSLKHPECSSTWSGKAIETHLAKQGRFPHDIK 240  
QY 241 PPLHAKNCDFSPFTGLQHTYDKIIMKKEKEBEGIEKGQILSSAADIATVQHTMACILYR 300  
DB 241 PPLHAKNCDFSPFTGLQHTYDKIIMKKEKEBEGIEKGQILSSAADIATVQHTMACILYR 300  
QY 276 GQILSSAADIATVQHTMACILYRTHRALPCKORDLPPONNAVLVASGVASNFYIR 335  
DB 301 GQILSSAADIATVQHTMACILYRTHRALPCKORDLPPONNAVLVASGVASNFYIR 360  
QY 336 ALEILTNAQCTLLCPPRLCTDNGIMIMWNGIERLRAGLGIHDIIEGIRYEPKPLGVD 395  
DB 361 ALEILTNAQCTLLCPPRLCTDNGIMIMWNGIERLRAGLGIHDIIEGIRYEPKPLGVD 420  
QY 396 ISKEVGEASIKVPOLKMEI 414  
DB 421 ISKEVGEASIKVPOLKMEI 439

RESULT 3  
Q96NH5 PRELIMINARY; PRT; 364 AA.  
AC Q96NH5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Hypothetical protein FLJ30879.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
RA Niinomiya K., Ishibashi T., Yamashita T., Murakawa K., Fujimori K.,  
RA Tanai H., Kimura M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Horuta T., Kusano J.,  
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,  
RA Togiyama S., Komai F., Hara R., Takeuchi K., Arita M., Inose N.,  
RA Musaeihino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshiikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujitani T.,  
RA Ono T., Yamada K., Hikiji T., Kobatake N., Inagaki S., Ikema Y., Okamoto S.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki S., Ikema Y., Okamoto S.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,



RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuko Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Iosugi T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 cDNAs.";  
 RL Nac. Genet. 36:40-45 (2004).  
 DR EMBL; AK055441; BAB70923.1; -.  
 DR MEROPS; M22.004; -.  
 DR GO; GO:0008450; P:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009095; Peptidase M22.  
 DR InterPro; IPR009180; Pept M22 Osa1gl.  
 DR Pfam; PF00814; Peptidase M22; 1.  
 DR PIRSF; PIRSF04537; Osa1glc\_ptide; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 KM Protease.  
 SQ SEQUENCE 364 AA; 39528 MW; E0B605A07D0EC3D6 CRC64;  
 Query Match 86.8%; Score 1845; DB 2; Length 364;  
 Best Local Similarity 99.2%; Pred. No. 8,4e-134;  
 Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MLILTKTAGVFPKSKRYEFLRSFNPFGTLFLHKIVLGLETSCDDTAAAVVDEITGV 60  
 DB 1 MLILTKTAGVFPKSKRYEFLRSFNPFGTLFLHKIVLGLETSCDDTAAAVVDEITGV 60  
 QY 61 LGEAIIHSTQVHLKTGGIVPPAAQOLHRENIORIVQBALASAGVSPSDLSAIAITIKPGL 120  
 DB 61 LGEAIIHSTQVHLKTGGIVPPAAQOLHRENIORIVQBALASAGVSPSDLSAIAITIKPGL 120  
 QY 121 ALSTGVLSFSLQVLGQKKPPIPIHMEAHALITRLTNKVEFPPLVLLISGHCILALV 180  
 DB 121 ALSTGVLSFSLQVLGQKKPPIPIHMEAHALITRLTNKVEFPPLVLLISGHCILALV 180  
 QY 121 ALSTGVLSFSLQVLGQKKPPIPIHMEAHALITRLTNKVEFPPLVLLISGHCILALV 180  
 DB 121 ALSTGVLSFSLQVLGQKKPPIPIHMEAHALITRLTNKVEFPPLVLLISGHCILALV 180  
 QY 181 QGVSDPFLIGKSLDIAPGMDLVKVARSLIHKPECSITMSGKAIIEHLAKQNRHFPDIK 240  
 DB 181 QGVSDPFLIGKSLDIAPGMDLVKVARSLIHKPECSITMSGKAIIEHLAKQNRHFPDIK 240  
 QY 241 PPLHAHAKNDPSFTGLQHTDITKIIMKEKEBEGIEKQILSSAADIAATVOHTMACHLVVR 300  
 DB 241 PPLHAHAKNDPSFTGLQHTDITKIIMKEKEBEGIEKQILSSAADIAATVOHTMACHLVVR 300  
 QY 301 THRAILFCCKQDILPONNAVIVASGVASNPYIRRALILITNAOTCTLLCPPRLCTDNG 360  
 DB 301 THRAILFCCKQDILPONNAVIVASGVASNPYIRRALILITNAOTCTLLCPPRLCTDNG 360  
 QY 361 IMIA 364  
 DB 361 IMIA 364  
 RESULT 4  
 Q6PEB4 PRELIMINARY; PRT; 414 AA.  
 AC Q6PEB4;  
 DT 05-JUL-2004 (Tremblere, 27, Created)  
 DT 05-JUL-2004 (Tremblere, 27, Last sequence update)  
 DT 05-JUL-2004 (Tremblere, 27, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCBI\_TaxID=10090;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH 11; TISSUE=Mammary tumor;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkin R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
 RA Diatchenko L., Manusina K., Farmer A.A., Rubin G.W., Hong L.,  
 RA Stetler M., Soares M.B., Bonaldo M.P., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
 RA Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH 11; TISSUE=Mammary tumor;  
 RA Strausberg R.;  
 RT Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC058172; AAH58172.1; -.  
 DR GO; GO:0008450; P:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009095; Peptidase M22.  
 DR InterPro; IPR009180; Pept M22 Osa1gl.  
 DR Pfam; PF00814; Peptidase M22; 1.  
 DR PIRSF; PIRSF04537; Osa1glc\_ptide; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBGAB CRC64;  
 Query Match 86.4%; Score 1835; DB 2; Length 414;  
 Best Local Similarity 85.0%; Pred. No. 5.8e-133;  
 Matches 352; Conservative 25; Mismatches 137; Indels 0; Gaps 0;

QY 1 MLILTKTAGVFPKSKRYEFLRSFNPFGTLFLHKIVLGLETSCDDTAAAVVDEITGV 60  
 DB 1 MLILTKTAGVFPKSKRYEFLRSFNPFGTLFLHKIVLGLETSCDDTAAAVVDEITGV 60  
 QY 61 LGEAIIHSTQVHLKTGGIVPPAAQOLHRENIORIVQBALASAGVSPSDLSAIAITIKPGL 120  
 DB 61 LGEAIIHSTQVHLKTGGIVPPAAQOLHRENIORIVQBALASAGVSPSDLSAIAITIKPGL 120  
 QY 121 ALSTGVLSFSLQVLGQKKPPIPIHMEAHALITRLTNKVEFPPLVLLISGHCILALV 180  
 DB 121 ALSTGVLSFSLQVLGQKKPPIPIHMEAHALITRLTNKVEFPPLVLLISGHCILALV 180  
 QY 121 ALSTGVLSFSLQVLGQKKPPIPIHMEAHALITRLTNKVEFPPLVLLISGHCILALV 180  
 DB 121 ALSTGVLSFSLQVLGQKKPPIPIHMEAHALITRLTNKVEFPPLVLLISGHCILALV 180  
 QY 181 QGVSDPFLIGKSLDIAPGMDLVKVARSLIHKPECSITMSGKAIIEHLAKQNRHFPDIK 240  
 DB 181 QGVSDPFLIGKSLDIAPGMDLVKVARSLIHKPECSITMSGKAIIEHLAKQNRHFPDIK 240  
 QY 241 PPLHAHAKNDPSFTGLQHTDITKIIMKEKEBEGIEKQILSSAADIAATVOHTMACHLVVR 300  
 DB 241 PPLHAHAKNDPSFTGLQHTDITKIIMKEKEBEGIEKQILSSAADIAATVOHTMACHLVVR 300  
 QY 301 THRAILFCCKQDILPONNAVIVASGVASNPYIRRALILITNAOTCTLLCPPRLCTDNG 360  
 DB 301 THRAILFCCKQDILPONNAVIVASGVASNPYIRRALILITNAOTCTLLCPPRLCTDNG 360  
 QY 361 IMIANGIERLPAAGIILHDIGIRYEPKPCGVNISKVGVASIVPOLKEI 414  
 DB 361 IMIANGIERLPAAGIILHDIGIRYEPKPCGVNISKVGVASIVPOLKEI 414  
 RESULT 5  
 Q8BLB6 PRELIMINARY; PRT; 414 AA.  
 ID Q8BLB6

AC Q8BLB6;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length  
 DE enriched library, clone:B230219017 product:similar to PUTATIVE  
 DE SIALOGLYCOPROTEASE TYPE 2.  
 GN Name=Osgpell1;  
 GN Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=21085560; PubMed=11217851; DOI=10.1038/35055500;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RT Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Komino H., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kanno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama T., Mishi K., Kitsuana T., Tashiro H., Itoh M.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hatada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hitozane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
 RA Kurihara C., Matuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito K., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AK045663; BAC32450.1; --

DR MEROPS; M22.004; -;  
 DR MG; MG1:191935; Osgpell1.  
 DR GO; GO:0008450; F-O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F-peptidase activity; IEA.  
 DR GO; GO:0008270; F-zinc ion binding; IEA.  
 DR GO; GO:0006508; Proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009095; Peptidase M22.  
 DR InterPro; IPR009180; Pept M22 Osa1gl.  
 DR Pfam; PF00814; Peptidase\_M22\_1; Osa1gl.  
 DR PIRSF; PIRSF004537; Osa1glc\_pptds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;  
 Query Match 86.0%; Score 1827; DB 2; Length 414;  
 Best Local Similarity 85.0%; Pred. No. 2,4e-132;  
 Matches 352; Conservative 23; Mismatches 39; Indels 0; Gaps 0;  
 QY 1 MLILTKAGVFPKPKSKRYEFLRSFNHPRGLFLHKIVLGIETSGDDPTAAAVDETGVN 60  
 DB 1 MLMLRTAGAIRPKPKSKYGFLLRRSVHPRTLSCRLVGLIETSCDDTGAAVDETGVN 60  
 QY 61 LGEALHSOTEVHLKTGIVPPAAQQLHRENIQIVQALASAGVSPSDLSAATIKPGL 120  
 DB 61 LGEALHSOTVHLKTGIVPPAAQQLHRENIQIVERTLSACRITPSDLSAATITIKPGL 120  
 QY 121 ALSLGVSFSLQVLQKPPPIPIHMEAAHLTLTNKVEPFLVLLISGCHLLALV 180  
 DB 121 ALSLGVSFSLQVLVQFKPFIPIHMEAAHLTLTNKVEPFLVLLISGCHLLALV 180  
 QY 181 QGVSDPFLLGKSLDIPAGMDLVARRSLIKPESTMSGKALSHLAKGNRFHPDIK 240  
 DB 181 QGVSDPFLLGKSLDIPAGMDLVARRSLIKPESTMSGKALSHLAKGNRFHPDIK 240  
 QY 241 PPLHAKNCDPFTGQVHTDKTIIMKEKEGIEKQIISADIAATVQHTVACHVVR 300  
 DB 241 PPLQNKNCDFSTGQHTDKITIKKEKEGIEKQIISADIAATVQHTVACHVVR 300  
 QY 301 TRAILFCQKQDLPPONNAVLAASGVASNFYIRALETITVNAOTCTLLCPPLCTDNG 360  
 DB 301 TRAILFCQKQDLPPONNAVLAASGVASNFYIRALETITVNAOTCTLLCPPLCTDNG 360  
 QY 361 IMIANNGIRLRAGLGIHDIGIRYEPKCPGVDSIKVGEASIVPOLKEI 414  
 DB 361 IMIANNGIRLRAGLGIHDIGIRYEPKCPGVDSIKVGEASIVPOLKEI 414  
 RESULT 6  
 ID Q9DONO PRELIMINARY; PRT; 414 AA.  
 AC Q9DONO;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length  
 DE enriched library, clone:2610001M19 product:similar to PUTATIVE  
 DE SIALOGLYCOPROTEASE TYPE 2.  
 GN Name=Osgpell1;  
 GN Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA STAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN PANTOM Consortium;  
 RL "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STAIN=C57BL/6J; TISSUE=Whole body;  
 RA The PANTOM Consortium;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/9r.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes."  
 RL Genome Res. 10:1617-1630(2000).  
 [5]  
 RP SEQUENCE FROM N.A.  
 RC STAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/9r.152600;  
 RA Shibata K., Itoh M., Alzawa K., Nagaoka S., Saeki N., Carninci P.,  
 RA Kono H., Akiyama J., Mishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Suni M., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer."  
 RL Genome Res. 10:1757-1771(2000).  
 [6]  
 RP SEQUENCE FROM N.A.  
 RC STAIN=C57BL/6J; TISSUE=Whole body;  
 RA Adachi J., Alzawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,  
 RA Imachi K., Ishii Y., Itoh M., Izawa M., Kaekawa T., Kato H.,  
 RA Kawai J., Kojima Y., Kono H., Konda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Nunazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito K., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami K., Tagawa A., Takahashi P., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK011265; BAB27506.1; --  
 DR MEROPS; M22.004; --  
 DR MGD; MGI:1919335; Osgp11.  
 DR GO; GO:00084450; P:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009095; Peptidase M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PRINTS; PRS004537; Osialglc\_ptds; 1.  
 DR PRINTS; PRS00789; OSIALOPTASB.  
 DR Prodom; PD002367; Peptidase\_M22; 1.  
 DR TRFPMAS; TIGR00329; gcp; 1.  
 KM Protease.  
 SQ SEQUENCE 414 AA; 44999 MW; 999BC689944DDB24 CRC64;  
 Query Match 85.6%; Score 1819; DB 2; Length 414;  
 Best Local Similarity 84.8%; Pctd No. 3; 5e-132;  
 Matches 351; Conservative 23; Mismatches 40; Indels 0; Gaps 0

[illegible]





RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun B.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Ceiniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champagne M., Dugan S.P., Friese E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,  
 RA Paollet J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
 RT "Finishing a whole-genome shotgun. Release 3 of the *Drosophila*  
 RT melanogaster euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Friese E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Ceiniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomic perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Betencourt B.R., Ceiniker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RG FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RG FlyBase;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AEO03513; AAF49008.1; -  
 DR Intact; O9VMD6; -  
 DR FlyBase; FBgn0031060; CG14231.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000905; Peptidase M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22\_1.  
 DR PRINTS; PRSF004537; Osialglc\_Pptds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TRIPRAMS; TIGR00329; GCP; 1.  
 DR SEQUENCE 409 AA; 45328 MW; 9797F66E7D155538 CRC64;  
 QY Query Match 33.6%; Score 714.5; DB 2; Length 409;  
 Best Local Similarity 42.0%; Pred. No. 1.3e-46;  
 Matches 156; Conservative 60; Mismatches 140; Indels 15; Gaps 6;

DB 27 VGIETSCDDTGAIVDTGRTVIANVLSQSFHTYGGIIPPRADLRARIESAYQRC 86  
 QY 99 LSASVSPEDLSAINTTKRGLASLGVLSLSLOVGLQKXKFTIHHMEANLIRLT 158  
 DB 87 MAAQOLKRPQLPAIVATTPRPGPLSLVLGVRPARHARLQKPLVHHMEADLDARME 146  
 QY 159 N-KYEPFLVLLIGSGHCLALVGVSPDLFLGSKSLDAPGMDLVARRSLIKHPEC 216  
 DB 147 HEPQIGYFPLCLASGGHCOLVAVANGPRLTLLGOTLDAPRPAFKIGRRRLRIHLPEY 206  
 QY 217 STMSGKAIIEHLAK-QGNRFHDIKPEPLHARNCDFSTGLQHVTDKIIMKEKEGIEK 275  
 DB 207 RLWNGGRAIEHNAQLASDPLAVEPFLPLAQQRNCNCSFAGIKNNSPRAIRAREARTPP 266  
 QY 276 GQILSSADIATVQHTMACHVKTRAILTC--KQRLPLPQNNANVYVASGVASNPFI 333  
 DB 267 DGVISNYGDFCAGLRSVSRHMRHRTORAIIEYCLPHRQLFGDPTPLVMSGVANNDDAI 326  
 QY 334 RRAELITVATQCTLLCPPEPLCTDNGIMIANNGIERLRAGGILHDIE-GIRYEPKCP 392  
 DB 327 YANTIELAAQYGRSFRPSKRYCSDNVMIAMHGVQL-----LDKKASTRYID--- 377  
 QY 393 GVDISKENGEA 403  
 DB 378 SIDIGSAGFA 388  
 RESULT 12  
 022145 PRELIMINARY; PRT; 480 AA.  
 ID 022145; OSVLM2;  
 AC 022145; 08VLM2;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Putative O-sialoglycoprotein endopeptidase (Sialoglycoprotease  
 DE GCP1).  
 GN Name:At2g45270; Synonyms=GCP1;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,  
 RA Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Town C.D., Kaul S.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Hausenhehl K., Adanaka I.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miyata M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,

RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narsaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.,  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC002387; AAB82636.2; -  
DR EMBL; AY024338; AAK00530.1; -  
DR EMBL; AY063864; AAL36220.1; -  
DR EMBL; AY117283; AAM51358.1; -  
DR PIR; E84888; E84888.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000905; Peptidase M22.  
DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
DR Pfam; PF00814; Peptidase\_M22; 1.  
DR PIRSF; PIRSF004537; Osialglc\_ptds; 1.  
DR PRINTS; PR00789; OSIALOPTASE.  
DR ProDom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
KW Protease.  
SQ SEQUENCE 480 AA; 52995 MW; 20DD6A86ACCIFFAD CRC64;  
Query Match 31.5%; Score 668.5; DB 2; Length 480;  
Best Local Similarity 39.4%; Pred. No. 5.4e-43;  
Matches 154; Conservative 60; Mismatches 130; Indels 47; Gaps 7;  
QY 38 IVLGIEIETSCDDTAAYVDEGTGNVLGEAIIHSQTEVHLKTGIVPPAAQQLHRENIQRIYQ 97  
DB 85 VILGIEIETSCDDTAAYVVRNGEIIISQVLSQAEILLVQGVAPKQAEHSHRVIDKVVQD 144  
QY 98 ALASAGVSPDLSAIIATTKKGLALSLGVGSLQVLGOLKPPPIPIHMEAAHLTR 157  
DB 145 ALDKANITKEDLSAIVAVYTIIGPGLSLCLRVGKARVAKRANSLPIYGVHMEAAHLVLR 204  
QY 158 T-NKVEPPLVLLISGGHCLTALVQVSDFLLGKSLDIAFGDMLDKVARRSLIKHPEC 216  
DB 205 VEGELSPFPMALLISGGHLLVLAHKLQYQLGTVDDALGEAFDKAKMGLDMH--- 261  
QY 217 STMSGKAIEHLAKQGNRFHFDIKPPLHAKNCDSPFTGLQHTVDTKIMKEKEGIEKG 276  
DB 262 --HSGGVAVEHLLEGDAKSKVFNVPKMYHDCNPSYAGLKTQVRLAIEAE---IDAK 315  
QY 277 QILSSA-----ADIAVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVLSGVA 328  
DB 316 CPVSSANEDRRNRADIIASFQVAVLHLEKCEAIDMALE---LPSIKMVISGVA 372  
QY 329 SNPIRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLIDIGIRYE- 387  
DB 373 SNKYVRLNNIVENKQLKLVCPSPSLCTDNGVAVAMTGLHFRVG-----RYDP 422  
QY 388 -----PKCPGLGVDSIKVGBA 403  
DB 423 PPATPEPDVYVLDPRMPLGEAYAKGRSEA 453  
RESULT 13  
Q73H71 PRELIMINARY; PRT; 335 AA.  
AC Q73H71;  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, last annotation update)  
DB Peptidase, M22 family protein.  
GN OrderedLocuNames=MD0659;  
OS Wolbachia pipiensis wmel;  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Wolbachiae; Wolbachia.  
OX NCBI\_TaxID=56077;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=15024419;

RA Wu M., Sun L.V., Yamathayan J.J., Riegler M., Deboy R.T.,  
RA Brownlie J.C., McGraw E.A., Martin W., Esser C., Ahmadijef N.,  
RA Wiegand C., Madupu R., Beanan M.J., Brinkac L.M., Daugherty S.C.,  
RA Durkin A.S., Kolonay J.F., Nelson W.C., Mohamoud Y., Lee P.,  
RA Berry K.J., Young M.B., Utterback T.R., Weidman J.F., Nierman W.C.,  
RA Paulsen I.T., Nelson K.E., Tettelin H., O'Neill S.L., Eisen J.A.;  
RT "Phylogenomics of the reproductive parasite Wolbachia pipiensis wmel:  
a streamlined genome overrun by mobile genetic elements.";  
RL Plos Biol. 2:327-341(2004).  
DR EMBL; AB017258; AAS14395.1; -  
DR TIGR; W00699; -  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR000905; Peptidase\_M22.  
DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
DR Pfam; PF00814; Peptidase\_M22; 1.  
DR PIRSF; PIRSF004537; Osialglc\_ptds; 1.  
DR PRINTS; PR00789; OSIALOPTASE.  
DR ProDom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
KW Complete proteome.  
SQ SEQUENCE 335 AA; 36634 MW; F4914CDA58BC9390 CRC64;  
Query Match 27.2%; Score 577.5; DB 2; Length 335;  
Best Local Similarity 36.3%; Pred. No. 3.6e-36;  
Matches 128; Conservative 68; Mismatches 130; Indels 27; Gaps 6;  
QY 37 KIVLGIEIETSCDDTAAYVDEGTGNVLGEAIIHSQTEVHLKTGIVPPAAQQLHRENIQRIYQ 96  
DB 2 KIIIAVETSCDEFAVAIVASDQVLAHEILSQAE-HKGGGVPIELASRAHMEHLSGLIK 60  
QY 97 EALSAGVSPDLSAIIATTKKGLALSLGVGSLQVLGOLKPPPIPIHMEAAHLTR 156  
DB 61 SAVERSNLNFCDLMAIATSGPGLIGLIVGTMAVAIAHVAKPPIAVNHLAAHLVIR 120  
QY 157 LNKVEPPLVLLISGGHCLTALVQVSDFLLGKSLDIAFGDMLDKVARRSLIKHPEC 216  
DB 121 LKHEVPEPPLVLLISGGHCLTALVQVSDFLLGKSLDIAFGDMLDKVARRSLIKHPEC 174  
QY 217 STMSGKAIEHLAKQGNRFHFDIKPPLHAKNCDSPFTGLQHTVDTKIMKEKEGIEKG 276  
DB 175 -SYPGPPLIKLAKQGNRFRFLPRAMIRSGCNPSFSGIKTVAKVLVQELKMS----- 228  
QY 277 QILSSAADIATVQHTMACHLVKTRTHAILFCQKORDLLPQNNNAVLSGVA SNFYIRA 336  
DB 229 ---QVDCVCA SFQECISIDILDRVSNAILMAESLINIKIND---FVITGVAAANNFLREK 282  
QY 337 LEILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLIDIGIRYE 389  
DB 283 LKQHN---LNIFFPNDICTNMAIMVGMWGIERLOKXN-----IDLPNAPR 327  
RESULT 14  
Q92LH8 PRELIMINARY; PRT; 360 AA.  
AC Q92LH8;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, last annotation update)  
DB PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).  
GN ORFNames=SWC03230;  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLIN=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Battu J.,  
RA Boistard P., Becker A., Boutry M., Cadieu B., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kist B., Lelaure V., Masny D.,







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 08:37:51 ; Search time 53.6902 Seconds  
(without alignments)  
1923.349 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHALTIRLTKNKEPPLV.....DISKEVGASIKVQLKMEI 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_16Dec04:\*

1: geneseqp1980s:\*\n2: geneseqp1990s:\*\n3: geneseqp2000s:\*\n4: geneseqp2001s:\*\n5: geneseqp2002s:\*\n6: geneseqp2003as:\*\n7: geneseqp2003bs:\*\n8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	267	5	ABG96487
2	1385	100.0	414	5	ABG96478
3	1385	100.0	414	5	ABG96478
4	1385	100.0	414	5	ABG96478
5	1348	97.3	414	5	ABG96478
6	1348	97.3	414	5	ABG96478
7	1348	97.3	414	5	ABG96478
8	1105	79.8	364	4	ADAB54771
9	415.5	30.0	409	4	ABG96478
10	409.5	29.6	463	3	AAV52216
11	398.5	28.8	245	3	AAV52216
12	398.5	28.8	439	3	AAV52216
13	398.5	28.8	444	3	AAV52216
14	342	24.7	251	5	AAE31054
15	321.5	23.2	312	6	ABU22934
16	314.5	22.7	312	6	ABU22934
17	308.5	22.3	348	6	ABU22934
18	307	22.2	350	8	ADL05040
19	304.5	22.0	343	6	ABU33303
20	291.5	21.0	341	4	AAU36205
21	291.5	21.0	341	6	ABU38276
22	291.5	21.0	241	7	ACG73342
23	291.5	21.0	401	7	ABO68626
24	289.5	20.9	341	7	ADG73344
25	289.5	20.9	342	3	AAV52202

26	289.5	20.9	342	4	AAU35450
27	289.5	20.9	342	6	ABU30280
28	289.5	20.3	340	6	ABU40514
29	280.5	20.3	357	7	ADP06228
30	279.5	20.2	337	4	AAU38187
31	279.5	20.2	337	6	ABU47561
32	279.5	20.2	337	6	ABU50237
33	277.5	20.0	335	6	ABU27480
34	277	20.0	309	6	ABU33614
35	276	20.0	341	6	ABU40069
36	276	19.9	421	5	ABG96491
37	275.5	19.9	337	4	AAU34711
38	275.5	19.9	337	6	ABU28771
39	270.5	19.5	325	2	AAV52203
40	270.5	19.5	325	3	AAV52203
41	270.5	19.5	337	3	AAV52204
42	268.5	19.4	343	7	ABO62704
43	266.5	19.2	338	6	ABM67812
44	263.5	19.0	354	6	ABU37844
45	263	19.0	341	6	ABU41689

#### ALIGNMENTS

RESULT 1	ABG96487	standard; protein, 267 AA.
ID	ABG96487	
AC	ABG96487	
XX		
DT	11-DEC-2002	(first entry)
XX		
DE	Novel human metalloprotease MPI fragment #1.	
XX		
KW	metalloprotease; MP-1; immune disorder; glutamate transport; cancer; motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes; reproductive disorder; Klinefelter's syndrome; germinal cell aplasia; genital wart; metabolic disorder; premature puberty; Kallman syndrome; Cushing's syndrome; neurodegenerative disease; Alzheimer's disease; Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis; liver disease; renal disease; immune disorder; rheumatoid arthritis; acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia; emphysema; cystic fibrosis; vascular disorder; inflammatory disorder; neurological disorder.	
KW		
OS	Homo sapiens.	
XX		
PN	WO200272751-A2.	
XX		
PD	19-SEP-2002.	
XX		
PF	05-FEB-2002; 2002WO-US003353.	
XX		
PR	05-FEB-2001; 2001US-0266519P.	
XX		
PR	10-APR-2001; 2001US-0282814P.	
XX		
PA	(BRIM ) BRISTOL-MYERS SQUIBB CO.	
XX		
PI	Chen J, Feder J, Nelson TC, Duclos F, Krystek S;	
XX		
DR	WPI, 2002-723329/78.	
XX		
DR	N-PSDB; ABS76639.	
XX		
PT	New isolated nucleic acid encoding MP-1 protein, useful for preventing, treating, or ameliorating diseases associated with aberrant metalloproteinase activity, e.g. immune, metabolic, inflammatory and neurological disorders.	
XX		
PT	Claim 5; Page 29; 473pp; English.	
XX		
CC	The invention describes an isolated nucleic acid molecule (I) encoding a metalloprotease (MP-1). (I) is useful for preventing, treating, or	

CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
CC sequence of a metalloprotease MPI protein  
XX  
SQ Sequence 267 AA;  
Query Match 100.0%; Score 1385; DB 5; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.4e-147;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEAHLTIRLTNKKVEPPLVLLISGHCILAVGVSDPLLGKSLDIAPGMDLDRVAR 60  
DB 1 MEAHLTIRLTNKKVEPPLVLLISGHCILAVGVSDPLLGKSLDIAPGMDLDRVAR 60  
QY 61 LSLIKHBCSTWSGKAIEHLAKQGNRFPDIKPLHAKNCDPFTGLQHTVDKTIIMKK 120  
DB 61 LSLIKHBCSTWSGKAIEHLAKQGNRFPDIKPLHAKNCDPFTGLQHTVDKTIIMKK 120  
QY 121 EKEEGIEKQILSSAADIAATVOHTMACLVKTRTHAILFCQKRDLLPNNNAVLSAGGV 180  
DB 121 EKEEGIEKQILSSAADIAATVOHTMACLVKTRTHAILFCQKRDLLPNNNAVLSAGGV 180  
QY 181 ASNFYIRRALBITLNATQCTLLCPPLRCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240  
DB 181 ASNFYIRRALBITLNATQCTLLCPPLRCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240  
QY 241 PKCPLGVDSIKVGEASIKVPOLKMEI 267  
DB 241 PKCPLGVDSIKVGEASIKVPOLKMEI 267  
DB 241 PKCPLGVDSIKVGEASIKVPOLKMEI 267  
RESULT 2  
ABG96478 standard; protein; 414 AA.  
XX  
AC ABG96478;  
XX  
DT 11-DEC-2002 (first entry)  
XX  
DE Novel human metalloprotease MPI.  
XX  
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
KW neurological disorder.  
XX  
XX Homo sapiens.  
XX  
OS Homo sapiens.  
XX  
PN MO200272751-A2.  
XX  
PD 19-SEP-2002.  
XX  
PF 05-FEB-2002; 2002WO-US003353.  
XX  
XX 05-FEB-2001; 2001US-0266518P.  
PR

PR 10-APR-2001; 2001US-0282814P.  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S,  
XX  
PI MPI; 2002-723329/78.  
DR N-PSDB; ABS76635.  
XX  
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.  
XX  
XX Claim 5; Fig 1A-C; 473pp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule (i) encoding a  
CC metalloprotease (MP-1). (i) is useful for preventing, treating, or  
CC ameliorating a medical condition, particularly an immune disorder, an  
CC aberrant glutamate transport or motor neuron disorder, such as  
CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
CC condition. The compositions and methods are also useful for diagnosing,  
CC prognosticating, treating, ameliorating and/or treating disorders  
CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
CC sequence of a metalloprotease MPI protein  
XX  
SQ Sequence 414 AA;  
Query Match 100.0%; Score 1385; DB 5; Length 414;  
Best Local Similarity 100.0%; Pred. No. 2.8e-147;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEAHLTIRLTNKKVEPPLVLLISGHCILAVGVSDPLLGKSLDIAPGMDLDRVAR 60  
DB 1 MEAHLTIRLTNKKVEPPLVLLISGHCILAVGVSDPLLGKSLDIAPGMDLDRVAR 207  
QY 61 LSLIKHBCSTWSGKAIEHLAKQGNRFPDIKPLHAKNCDPFTGLQHTVDKTIIMKK 120  
DB 208 LSLIKHBCSTWSGKAIEHLAKQGNRFPDIKPLHAKNCDPFTGLQHTVDKTIIMKK 267  
QY 121 EKEEGIEKQILSSAADIAATVOHTMACLVKTRTHAILFCQKRDLLPNNNAVLSAGGV 180  
DB 268 EKEEGIEKQILSSAADIAATVOHTMACLVKTRTHAILFCQKRDLLPNNNAVLSAGGV 327  
QY 181 ASNFYIRRALBITLNATQCTLLCPPLRCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240  
DB 328 ASNFYIRRALBITLNATQCTLLCPPLRCTDNGIMIANNGIERLRAGLILHIDIGIRYE 387  
QY 241 PKCPLGVDSIKVGEASIKVPOLKMEI 267  
DB 388 PKCPLGVDSIKVGEASIKVPOLKMEI 414  
RESULT 3  
ABB05481 standard; protein; 414 AA.  
XX  
XX ABB05481;  
XX  
DT 19-APR-2002 (first entry)  
XX  
DE Human O-sialoglycoproteinase-like protein SEQ ID NO:2.  
XX  
XX Human; O-sialoglycoproteinase-like protein; OSGLP; enzyme.  
XX

OS Homo sapiens.  
 XX CN1318550-A.  
 XX  
 PD 24-OCT-2001.  
 XX  
 PF 19-APR-2000; 2000CN-00106834.  
 XX  
 PR 19-APR-2000; 2000CN-00106834.  
 XX  
 PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.  
 XX  
 PI Mao Y, Xie Y;  
 XX  
 DR WPI: 2002-115090/16.  
 DR N-PSDB; ABA93268.  
 XX  
 PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful  
 XX for diagnosing, preventing and treating related diseases.  
 XX  
 PS Claim 1; Page 30-31 (Disclosure); 38pp; Chinese.

CC The present sequence represents human O-sialoglycoproteinase-like protein  
 CC (OSGPRP). The present invention also describes: (1) the preparation of  
 CC the OSGPRP protein; (2) applying the OSGPRP protein in diagnosis; (3) the  
 CC prevention and/or treatment of related diseases; (4) utilizing the OSGPRP  
 CC protein in screening its agonist, excitomotor and inhibitor and preparing  
 CC an antibody against the OSGPRP protein; and (5) the use of the OSGPRP  
 CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors  
 CC and antibodies in treating diseases related to the abnormal OSGPRP gene  
 CC and in preparing the medicine composite for the treatment

XX  
 SQ Sequence 414 AA;

Query Match 100.0%; Score 1385; DB 5; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-147;  
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAALATRLTNKVEPPVLVLLISGHCILALVQGVSPFLIGKSLDIAFGMDLVARR 60  
 DB 148 MEAALATRLTNKVEPPVLVLLISGHCILALVQGVSPFLIGKSLDIAFGMDLVARR 207  
 QY 61 LSLIKHEPCSTMSGKATIEHLAKQGNRFHFDIKPLHAKNDCPSFTGQHTYTDKIMKK 120  
 DB 208 LSLIKHEPCSTMSGKATIEHLAKQGNRFHFDIKPLHAKNDCPSFTGQHTYTDKIMKK 267  
 QY 121 EKEGIEKGQILSSAADIATVQHTMACHLVKTRTALIFCKQBDLPQNNAVLVASGV 180  
 DB 268 EKEGIEKGQILSSAADIATVQHTMACHLVKTRTALIFCKQBDLPQNNAVLVASGV 327  
 QY 181 ASNFYIRRALILTNATQCTLLCPPLCTDNGMIANNGIERLRAGLIHDIGIRYE 240  
 DB 328 ASNFYIRRALILTNATQCTLLCPPLCTDNGMIANNGIERLRAGLIHDIGIRYE 387  
 QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267  
 DB 388 PKCPLGVDISKVEGASIKVPOLKMEI 414

RESULT 4  
 ABJ26654  
 ID ABJ26654 standard; protein; 414 AA.  
 XX  
 AC ABJ26654;  
 XX  
 DT 01-MAY-2003 (first entry)

DE Human protein modification + maintenance molecule protein SEQ ID No. 8.  
 XX  
 KM Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
 KM cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
 KM antitumor; hepatotropic; gynaecological; antibacterial; virucide;  
 KM protozoacide; antiparasitic; cell proliferative disease; PMOD;

KW protein modification and maintenance molecule; immunogenic fragment;  
 KW cancer; autoimmune; inflammatory disease; neurological disorder;  
 KW gastrointestinal; developmental; vesicle trafficking disorder; infection;  
 KW protein-protein interaction; drug-target interaction;  
 KW gene expression profile; human.

OS Homo sapiens.  
 XX  
 OS  
 XX  
 PN WO200300844-A2.  
 XX  
 PD 03-JAN-2003.  
 XX  
 PF 18-JUN-2002; 2002WO-US019360.  
 XX  
 PR 22-JUN-2001; 2001US-0300508P.  
 XX  
 PR 06-JUL-2001; 2001US-0303445P.  
 PR 13-JUL-2001; 2001US-0303405P.  
 PR 09-AUG-2001; 2001US-0311442P.  
 PR 24-AUG-2001; 2001US-0314821P.  
 PR 29-AUG-2001; 2001US-0315929P.  
 PR 03-MAY-2002; 2002US-0378205P.

XX (INCY-) INCYTE GENOMICS INC.  
 XX  
 PA Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BM;  
 PI Warren BA, Ison CH, Honchell CD, Nguyen DB, Lu DM, Lee BA, Yue H;  
 PI Forgythe LJ, Barroso I, Ramkumar J, Griffin JA, Li JX, Yang J;  
 PI Thangavelu K, Gietzen KJ, Ding L, Baugman MR, Borowsky ML, Yao MG;  
 PI Walla NK, Mason PM, Gururajan R, Lee S, Becha SD, Lee SY, Tran UK;  
 PI Elliott VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebajadian Y;  
 XX  
 DR WPI: 2003-184039/18.  
 DR N-PSDB; AET23207.

PT New isolated human PMOD polypeptide and polynucleotide, useful for  
 PT diagnosing, treating and preventing diseases or conditions associated  
 PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  
 PT infections.  
 XX  
 PS Claim 63; Page 182-193; 225pp; English.

CC The invention relates to an isolated polypeptide comprising: any of 28  
 CC sequences of 48-1256 amino acids; a natural amino acid sequence at least  
 CC 90% identical to the 28 amino acid sequences; 94% identical to a sequence  
 CC of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
 CC acids, or 97% identical to a sequence of 242 amino acids, all given in  
 CC the specification; or a biologically active or immunogenic fragment of  
 CC the isolated polypeptide. The polypeptides and polynucleotides are useful  
 CC in diagnosing, treating and preventing diseases or conditions associated  
 CC with the decreased expression of protein modification and maintenance  
 CC molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
 CC atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
 CC allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
 CC epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
 CC endometriosis), developmental, vesicle trafficking disorders, and  
 CC infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
 CC useful in assessing the effects of exogenous compounds on the expression  
 CC of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
 CC fragments are useful in screening compounds for effectiveness as agonist  
 CC or antagonist of the polypeptides, or in altering the expression of the  
 CC target polynucleotide and compounds that specifically bind to or modulate  
 CC the activity of the polypeptide. The microarray is useful in monitoring  
 CC or measuring protein-protein interactions, drug-target interactions, and  
 CC gene expression profiles. This sequence represents a human PMOD protein  
 CC of the invention

SQ Sequence 414 AA;

Query Match 100.0%; Score 1385; DB 5; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-147;  
 Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAALATRLTNKVEPPVLVLLISGHCILALVQGVSPFLIGKSLDIAFGMDLVARR 60

```

Db      148 MEAAHLLTIRLTNNKVFPPVLVLLISGGHCLALVQGVSDPLLKGSIDIAFGMDLVARR 207
Qy      61 LSLIKHPECSMTSGGKAIEHLAKQGNRFPHDIKPLPHAKNDDPSTGLOHTVDKIIMK 120
Db      208 LSLIKHPECSMTSGGKAIEHLAKQGNRFPHDIKPLPHAKNDDPSTGLOHTVDKIIMK 267
Qy      121 EKEBGIEKGQILSSAADIAATVQHTMACLVKTRTRAILFCQKORDLLPOMNAVLSAGGV 180
Db      268 EKEBGIEKGQILSSAADIAATVQHTMACLVKTRTRAILFCQKORDLLPOMNAVLSAGGV 327
Qy      181 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHIDIGIRYE 240
Db      328 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHIDIGIRYE 387
Qy      241 PKCPILGVNISKVGEASIKVPOLKMEI 267
Db      388 PKCPILGVNISKVGEASIKVPOLKMEI 414

```

## RESULT 5

AAE29234

ID AAE29234 standard; protein; 414 AA.

XX AAE29234;

XX 27-JAN-2003 (first entry)

XX Human glycoprotease 28472 protein.

Human; adenosine deaminase; seven transmembrane domain receptor; cancer;  
 KM 7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arthralgia;  
 KM rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;  
 KM hyperension; ischaemic heart disease; obesity; myocardial infarction;  
 KM endothelial cell disorder; Grave's disease; psoriasis; brain disorder;  
 KM Parkinson's disease; Alzheimer's disease; haematopoietic disorder;  
 KM cerebral oedema; metabolic disorder; liver disorder; platelet disorder;  
 KM chromosome mapping; tissue typing; gene therapy; neuroprotective;  
 KM cyrostatic; anorectic; cardiant; haemostatic.

XX Homo sapiens.

XX Key Location/Qualifiers

XX 1. 108 /note= "Non-transmembrane domain; N-terminal cytoplasmic domain"

XX 38. 369 /note= "Endopeptidase O-sialoglycoprotein hydrolase"

XX 148 MEAAHLLTIRLTNNKVFPPVLVLLISGGHCLALVQGVSDPLLKGSIDIAFGMDLVARR 207

XX 61 LSLIKHPECSMTSGGKAIEHLAKQGNRFPHDIKPLPHAKNDDPSTGLOHTVDKIIMK 120

XX 208 LSLIKHPECSMTSGGKAIEHLAKQGNRFPHDIKPLPHAKNDDPSTGLOHTVDKIIMK 267

XX 121 EKEBGIEKGQILSSAADIAATVQHTMACLVKTRTRAILFCQKORDLLPOMNAVLSAGGV 180

XX 268 EKEBGIEKGQILSSAADIAATVQHTMACLVKTRTRAILFCQKORDLLPOMNAVLSAGGV 327

XX 181 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHIDIGIRYE 240

XX 328 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHIDIGIRYE 387

XX 241 PKCPILGVNISKVGEASIKVPOLKMEI 267

XX 388 PKCPILGVNISKVGEASIKVPOLKMEI 414

XX Domain

XX 109. 132 /note= "Transmembrane domain"

XX 133. 164 /note= "Non-transmembrane domain; non-cytoplasmic loop"

XX 138. 152 /note= "Glycoprotease domain"

XX 165. 189 /note= "Transmembrane domain"

XX 190. 316 /note= "Transmembrane domain"

XX 317. 333 /note= "Non-transmembrane domain; cytoplasmic domain"

XX 334. 414 /note= "Transmembrane domain"

XX 374. 414 /note= "Non-transmembrane domain"

XX /note= "Sialoglycoprotease type domain"

XX W0200274960-A2.

XX 26-SEP-2002.

XX 08-NOV-2001; 2001WO-US051427.

PR 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 PA (MILL-) MILLENNIUM PHARM INC.  
 PI Leiby KR, Kapeller-Libermann R, Glucksmann M;  
 DR MPI, 2002-759898/82.  
 DR N-PSDB; AAD46856.  
 XX New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,  
 PT useful for diagnosing and treating cancer, immune, cardiovascular,  
 PT hemotopoietic, brain, pain, metabolic, liver or platelet disorders, and  
 PT in pharmacogenomics.  
 XX Claim 1; Fig 8; 178pp; English.

XX The present invention relates to novel 38650, 28472, 5495, 65507, 81588  
 CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-  
 CC protease or seven transmembrane domain (7TM) receptor family members.  
 CC Sequences of the invention are useful in diagnosing and treating cancer  
 CC or aberrant cellular proliferation and/or differentiation (e.g. colon or  
 CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid  
 CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,  
 CC hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,  
 CC myocardial infarction, thrombus) including endothelial cell disorders  
 CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain  
 CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),  
 CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet  
 CC disorders. They are also useful in screening assays, predictive medicine  
 CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials  
 CC and pharmacogenetics) and prophylactic and therapeutic methods. The  
 CC nucleic acids may also be used in chromosome mapping, tissue typing and  
 CC forensic biology and as surrogate markers. Sequences of the invention are  
 CC also used in gene therapy. The present sequence is human glycoprotease  
 CC 28472 protein

XX Sequence 414 AA;

XX Query Match 97.3%; Score 1348; DB 5; Length 414;

XX Best Local Similarity 97.4%; Pred. No. 4.3e-143;

XX Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

XX 1 MEAAHLLTIRLTNNKVFPPVLVLLISGGHCLALVQGVSDPLLKGSIDIAFGMDLVARR 60

XX 148 MEAAHLLTIRLTNNKVFPPVLVLLISGGHCLALVQGVSDPLLKGSIDIAFGMDLVARR 207

XX 61 LSLIKHPECSMTSGGKAIEHLAKQGNRFPHDIKPLPHAKNDDPSTGLOHTVDKIIMK 120

XX 208 LSLIKHPECSMTSGGKAIEHLAKQGNRFPHDIKPLPHAKNDDPSTGLOHTVDKIIMK 267

XX 121 EKEBGIEKGQILSSAADIAATVQHTMACLVKTRTRAILFCQKORDLLPOMNAVLSAGGV 180

XX 268 EKEBGIEKGQILSSAADIAATVQHTMACLVKTRTRAILFCQKORDLLPOMNAVLSAGGV 327

XX 181 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHIDIGIRYE 240

XX 328 ASNFYIRRALILTNATQCTLLCPPRLCTDNGIMIANWGIERLRAGLGIHIDIGIRYE 387

XX 241 PKCPILGVNISKVGEASIKVPOLKMEI 267

XX 388 PKCPILGVNISKVGEASIKVPOLKMEI 414

XX RESULT 6

XX ID ABG71161 standard; protein; 414 AA.

XX AC ABG71161;

XX 30-JAN-2003 (first entry)

XX DT

XX

XX

DE Novel human glycoprotease 28472.  
 XX  
 KM Cancer; aberrant cell proliferation; aberrant cell differentiation;  
 KM breast cancer; ovarian cancer; prostate cancer; colon cancer;  
 KM lung cancer; immune disorder; heart disorder; cardiovascular disorder;  
 KM endothelial disorder; hematopoietic disorder; blood vessel disorder;  
 KM brain disorder; pain; metabolic disorder; liver disorder; diabetes;  
 KM platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;  
 KM autoimmune disorder; hypertension; atherosclerosis; heart failure;  
 KM myocardial infarction; ischaemic heart disease; Crohn's disease;  
 KM Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;  
 KM cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;  
 KM Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.  
 OS Homo sapiens.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT 138. .152  
 FT Domain /label=Glycoprotease\_domain  
 XX  
 PN MO20027233-A2.  
 XX  
 XX 03-OCT-2002.  
 PD  
 PF 08-NOV-2001; 2001WO-US046724.  
 XX  
 PF 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Lei by KR, Kapeller-Libermann R, Glucksmann M;  
 XX  
 DR WPI; 2003-029938/02.  
 DR N-PSDB; ABS57020.  
 XX  
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain  
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,  
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or  
 PT hypertension.  
 XX  
 PS Claim 4; Fig 8A-B; 178pp; English.  
 XX  
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and  
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The  
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine  
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding  
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise  
 CC sequences that encode a human seven transmembrane domain (7TM). The  
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide  
 CC sequences are useful for diagnosing, preventing or treating a subject  
 CC with or at risk of developing a disorder, e.g. cancer or aberrant  
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,  
 CC prostate, colon or lung cancer), immune disorders, heart disorders,  
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,  
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,  
 CC liver disorders or platelet disorders. These disorders include carcinoma,  
 CC sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,  
 CC hypertension, atherosclerosis, heart failure, myocardial infarction,  
 CC ischaemic heart disease, Crohn's disease, Grave's disease, Kawasaki  
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral  
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,  
 CC cachexia or diabetes. This is the amino acid sequence of the novel human  
 CC glycoprotease 28472  
 XX  
 XX Sequence 414 AA:  
 Query Match 97.3%; Score 1348; DB 6; Length 414;  
 Best Local Similarity 97.4%; Freq. No. 4.3e-143;  
 Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

DB 148 MEAHALTIRLTNKKVEPPVLVLLISGGHCLALVQGVSDFLICKSLDIAPGMDLVARR 207  
 QY 61 LSLIKPECSSTSGGATIHAKQGRFRFPDIKPRILHAKNCFSTGLOHTDKIMKK 120  
 DB 208 LSLIKPECSSTSGGATIHAKQGRFRFPDIKPRILHAKNCFSTGLOHTDKIMKK 267  
 QY 121 EKEEGIEKQIISADIAATVQHTMACILVYKTRTHAILFCRKORDLLPQNNAVLVASGGV 180  
 DB 268 KQBEIEKQIISADIAATVQHTMACILVYKTRTHAILFCRKORDLLPQNNAVLVASGGV 327  
 QY 181 ASNFYIRRALBITLVNATQCTLLCPPELCTDNGIMIMWNGIERLRAGLGIHDIEGIRYE 240  
 DB 328 ASNFYIRRALBITLVNATQCTLLCPPELCTDNGIMIMWNGIERLRAGLGIHDIEGIRYE 387  
 QY 241 PKCPILGVDISKVEGASIVPOLKMEI 267  
 DB 388 PKCPILGVDISKVEGASIVPOLKMEI 414  
 RESULT 7  
 AB009569  
 ID AB009569 standard; protein; 414 AA.  
 XX  
 AC AB009569;  
 XX  
 DT 08-JUL-2003 (first entry)  
 XX  
 DE Human glycoprotease encoded by cDNA 28472.  
 XX  
 KM Human; enzyme; cancer; aberrant cellular proliferation; differentiation;  
 KM immune disorders; heart disorder; brain disorder;  
 KM cardiovascular disorder; endothelial cell disorder; pain disorder;  
 KM hematopoietic disorder; blood vessel disorder; metabolic disorder;  
 KM liver disorder; platelet disorder; glycoprotease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003009017-A1.  
 PN  
 XX 09-JAN-2003.  
 PD  
 PF 08-NOV-2001; 2001US-00012140.  
 XX  
 PF 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX  
 PA (LEIBY) LEIBY K R.  
 PA (KAPL) KAPPELLER-LIBERMANN R.  
 PA (GLUC/) GLUCKSMANN M A.  
 XX  
 PI Lei by KR, Kapeller-Libermann R, Glucksmann M;  
 XX  
 DR WPI; 2003-428868/40.  
 DR N-PSDB; ACA60887.  
 XX  
 PT New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
 PT molecules, useful for diagnosing, treating cancer, pain, or immune,  
 PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic  
 PT and liver disorders.  
 XX  
 PS Claim 4; Fig 8; 90pp; English.  
 XX  
 CC The invention relates to an isolated 38650 (encoding adenosine  
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
 CC a sequence which is at least 60% identical to the six nucleic acids or  
 CC their open reading frames, fragments of at least 15 nucleotides,  
 CC naturally occurring variants, or a DNA insert of the plasmid deposited  
 CC with the American Type Culture Collection as Accession No. not defined in  
 CC the specification, which encodes the amino acid sequence). Also included  
 CC are a host cell containing the nucleic acids (used to produce the

CC proteins), the encoded proteins, an antibody that selectively binds to  
 CC the polypeptide, and identifying a compound that binds to/modulates the  
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
 CC methods are useful for diagnosing, treating cancer, aberrant cellular  
 CC proliferation and/or differentiation, immune disorders, heart disorders,  
 CC cardiovascular disorders including endothelial cell disorders,  
 CC hematopoietic disorders, blood vessel disorders, brain disorders, pain  
 CC and metabolic disorders, liver disorders and platelet disorders (many  
 CC examples of these disorders are given in the specification). The present  
 CC sequence is the Human glycoprotease encoded by cDNA 28472

XX Sequence 414 AA;

Query Match 97.3%; Score 1348; DB 6; Length 414;

Best Local Similarity 97.4%; Pred. No. 4.3e-143; Indels 0; Gaps 0;

Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MEAHALTIRLTNNKVEFPFLVLLISGHCILALVQGVSDFLLGKSLDIAPGMLDKVAR 60  
 DB 148 MEAHALTIRLTNNKVEFPFLVLLISGHCILALVQGVSDFLLGKSLDIAPGMLDKVAR 207  
 QY 61 LSLIKHPECSITMSGKAIIEHLAKQGNRFHDIKPLHAKNCDPSFTGLQHTYTDKTIIMKX 120  
 DB 208 LSLIKHPECSITMSGKAIIEHLAKQGNRFHDIKPLHAKNCDPSFTGLQHTYTDKTIIMKX 267  
 QY 121 EKEBIEKQIILSSADIAATVQHTMACHLVYRTRAILFCRORDLLPNNNAVLSAGGV 180  
 DB 268 EKEBIEKQIILSSADIAATVQHTMACHLVYRTRAILFCRORDLLPNNNAVLSAGGV 327  
 QY 181 ASNFYIRALIEITNATQCTLLCPPLCTDNGIMIANNGIERLADGLIDIGIRYE 240  
 DB 328 ASNFYIRALIEITNATQCTLLCPPLCTDNGIMIANNGIERLADGLIDIGIRYE 387  
 QY 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267  
 DB 388 PKCPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 8

ADA54471 ID ADA54471 standard; protein; 364 AA.

XX ADA54471;

XX 20-NOV-2003 (first entry)

XX Human protein, SEQ ID 2039.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nocotropic;

KW Gene Therapy; human; secretory protein; membrane proteins; cancer;

KW inflammatory disease; osteoporosis; neurological disease.

XX Homo sapiens.

XX EP1293569-A2.

XX 19-MAR-2003.

XX 21-MAR-2002; 2002EP-00006586.

XX 14-SEP-2001; 2001JP-00328381.

XX 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.

XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamechika I;

XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maehno Y;

XX WPI; 2003-395539/38.

XX N-PSDB; ADA52832.

PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.

XX Claim 14; SEQ ID NO 2039; 205pp; English.

PS The present invention relates to novel human secretory or membrane  
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.

XX Sequence 364 AA;

Query Match 79.8%; Score 1105; DB 6; Length 364;

Best Local Similarity 98.6%; Pred. No. 1.1e-115; Indels 0; Gaps 0;

Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MEAHALTIRLTNNKVEFPFLVLLISGHCILALVQGVSDFLLGKSLDIAPGMLDKVAR 60  
 DB 148 MEAHALTIRLTNNKVEFPFLVLLISGHCILALVQGVSDFLLGKSLDIAPGMLDKVAR 207  
 QY 61 LSLIKHPECSITMSGKAIIEHLAKQGNRFHDIKPLHAKNCDPSFTGLQHTYTDKTIIMKX 120  
 DB 208 LSLIKHPECSITMSGKAIIEHLAKQGNRFHDIKPLHAKNCDPSFTGLQHTYTDKTIIMKX 267  
 QY 121 EKEBIEKQIILSSADIAATVQHTMACHLVYRTRAILFCRORDLLPNNNAVLSAGGV 180  
 DB 268 EKEBIEKQIILSSADIAATVQHTMACHLVYRTRAILFCRORDLLPNNNAVLSAGGV 327  
 QY 181 ASNFYIRALIEITNATQCTLLCPPLCTDNGIMIA 217  
 DB 328 ASNFYIRALIEITNATQCTLLCPPLCTDNGIMIA 364

RESULT 9

ABB69133 ID ABB69133 standard; protein; 409 AA.

XX ABB69133;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 34191.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL13236.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

XX Disclosure; SEQ ID NO 34191; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 409 AA;

Query Match 30.0%; Score 415.5; DB 4; Length 409;

Best Local Similarity 38.2%; Pred. No. 1.4e-37;

Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;

QY 1 MEAHLTIRLTN--KVEPFLVLLISGGHCLLAVQVSDPFLLGKSLDIAPGMDLDRVA 58  
 DB 136 MEAHLQARMEHPQIGYPLCLLASGGHCOLVANNGGRITLLGQTLDDAPGEAFDKIG 195

QY 59 RLSLIHPECSTMSGKAI EHLAK--QGNRPHFDIKPRLHAKNCDFEFTGLQHTDKIT 117  
 DB 196 RLRLHLTPERLWNGRAIEHAQLASDPLAEPFLPLAQRNCFSPAGIKNNSPFAI 255

QY 118 MKKEKEGIEKGQILSSADIAATVQHTMACHLVYKTRRALTFC--KQRDLDPQNNAVLV 175  
 DB 256 RAERBARTRPPDGVISYNGDFCAGILASVSHLMKRTIRALRYCLLPHRQLPGDTPPLTV 315

QY 176 ASGGVSNFYIRALEITLTNATQCTLLCPPRLCTDNGIMIANGIERTLRAGLGIHLDIE 235  
 DB 316 MSGGVANNDAIYANIEHLAAQYGRSPRSKRYCSDNVMIAMNGVEGL-----LDQKE 369

QY 236 -GIRYEKCPGVDISKVEGEA 256  
 DB 370 ASTRYDYD---SIDIQSAGFA 388

RESULT 10

AAV52216 standard; protein; 463 AA.

XX AAV52216; 2

XX 09-FEB-2000 (first entry)

XX Arabidopsis thaliana yjld protein homologue.

XX yjld protein: essential; Gram positive; Gram negative; conserved; motif;  
 KW identification; antagonist; antibacterial; antibiotic; broad spectrum;  
 KM treatment; infection; resistance; drug target.

OS Arabidopsis thaliana.

XX Key Location/Qualifiers

FT Region 86..96 /note="yjld conserved motif 3"

FT Region 111..131 /note="yjld conserved motif 4"

FT Region 152..198 /note="yjld conserved motif 2"

FT Region 208..259 /note="yjld conserved motif 1"

XX MO9954470-A2.

XX 28-OCT-1999.

XX 20-APR-1999; 99WO-BP002635.

XX 22-APR-1999; 99GB-00009423.

XX (GLAXO ) GLAXO GROUP LTD.

PI Arizoni F, Edgerton MD, Loferer H, Peltsch MC,  
 XX WPI, 2000-013253/01.  
 DR Novel bacterial polypeptides used to identify broad spectrum antibiotics.  
 XX Claim 1; Fig 1; 55pp; English.

CC Proteins AAV52202-Y52217 encompass a novel family of proteins designated  
 CC the yjld family, after the name given to the *Escherichia coli* family  
 CC member. These proteins are essential for the survival of both Gram  
 CC negative and Gram positive bacteria, although no function has as yet been  
 CC ascribed to these proteins. The yjld proteins, fragments of yjld proteins  
 CC (for example, fragments encompassing one or more conserved yjld motifs  
 CC such as AAV52218-Y52284) and nucleotides encoding them can be used to  
 CC identify antagonists and broad spectrum antibacterial compounds. These  
 CC antagonists and compounds can be used to treat a wide range of bacterial  
 CC infections. New antibiotics are urgently needed, as serious bacterial  
 CC infections and antibiotic resistant strains are becoming increasingly  
 CC prevalent. The proteins of the invention are essential proteins for  
 CC bacterial viability, and represent new targets for antibiotics

XX Sequence 463 AA;

Query Match 29.6%; Score 409.5; DB 3; Length 463;

Best Local Similarity 36.6%; Pred. No. 8.2e-37;

Matches 100; Conservative 37; Mismatches 91; Indels 45; Gaps 6;

QY 1 MEAHLTIRLTN--KVEPFLVLLISGGHCLLAVQVSDPFLLGKSLDIAPGMDLDRVA 59  
 DB 192 MEAHLVLRVQLBELSPFMALLISGHNLLVLAHKLQGYTGLGTVDAIIEAPDKTK 251

QY 60 RLSLIHPECSTMSGKAI EHLAK--QGNRPHFDIKPRLHAKNCDFEFTGLQHTDKIT 119  
 DB 252 WLGLDMH-----RSGGPAVEBELALEGDAKSVFNPMPMKHKNCFSPAGIKNNSPFAI 306

QY 120 KEKEGIEKGQILSSADIAATVQHTMACHLVYKTRRALTFC--KQRDLDPQNNAVLV 179  
 DB 307 KE-----IRRADIASFORVAVLHLEKGERALDMLR--LSPSIKHWYISGG 353

QY 180 VASNFYIRALEITLTNATQCTLLCPPRLCTDNGIMIANGIERTLRAGLGIHLDIEGIRY 239  
 DB 354 VASNRYVRLRLNNIYENKRLKLVCPPLSLCTDNGVAVMTGLEHPRVG-----RY 403

QY 240 E-----KCPPLGVDISKVEGEA 256  
 DB 404 DPPPATPEPDVYDLRPRMPLGEBYAKGRSBA 436

RESULT 11

AAG19287 standard; protein; 245 AA.

XX AAG19287;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 21031.

XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KM termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000BP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-013548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142399P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.

PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 24-SEP-1999; 99US-0156458P.  
PR 28-SEP-1999; 99US-0156596P.  
PR 29-SEP-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.



PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158023P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 28.8%; Score 398.5; DB 3; Length 245;  
Best Local Similarity 38.7%; Pred. No. 5.6e-36;  
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

QY 1 MEAAATLIRLT-NKVEPFLVLLISGGCHLALVGVSDPFLIGSLDIAPQMDIKAR 59  
DB 1 MEAAATLIRLT-NKVEPFLVLLISGGCHLALVGVSDPFLIGSLDIAPQMDIKAR 60  
QY 60 RLSTIKPECSITMSGKAIIEHLAKQGNRFHDIKPELHANNKDPSPFGLQHVTDKIMK 119  
DB 61 WLGLDMH-----RSGGPAVEBELAEGDAKSKYKFNPMKYHDCNNSYGLKQVRLAIEA 115  
QY 120 KEKEEGIEKQQLISA-----ADIAATVOHTWACHLVKRTYRAILFCQQRDLLPQNN 171  
DB 116 KE-----IDAKCPVSSATVEDRNRADIAASFQVAVLHBEKCEAIOWALE---LBPSI 168  
QY 172 AVLVASGCVASNFYIRALLETILTAATCTLLCPPRRLCTDNGIMTANNGIERLAGLIL 231  
DB 169 KHWISGCVASNKYVRLNLNIVENKNILKVCPPPSLCTDNGVVAWVAGLBEHFRVG----- 224  
QY 232 HDIEGIRYEPKCP 244  
DB 225 -----RYDPPP 231

RESULT 12  
AAAG19286  
ID AAAG19286 standard; protein: 439 AA.

XX AC AAAG19286;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21030.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 17-JUN-1999; 99US-0139494P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.

PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154799P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157533P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159355P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161932P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 28.8%; Score 398.5; DB 3; Length 439;  
Best Local Similarity 38.7%; Pred. No. 1,3e-35;  
Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

QY 1 MEAAATIRLT-NKVEPPPLVLLISGCHLALVOGVSDPFLLLGSLDIAPGDMLDKVAR 59  
DB 195 MEAHTLVARLVQELSPFMALLISGHNVLVIAHKLGQYTGTVDDAIGBAFDKTA 254  
QY 60 RLSLKHPKPCSTMSGKALIEHLAKOGRHFPDKPPLHNAKNCDSFTGLQHTDKIIMK 119  
DB 255 WLGLDMH-----RSGGPAVEELALBGDAKSVFNPMPKXKCNFVAGIKTOVRLAIFA 309  
QY 120 KEKEGIEKQILISA-----ADIAATVOHTMACHLVKTRHRIILFCOKRDLLPNN 171  
DB 310 KE-----IDAKCPVSSATNEDRRNRADIASFQRAVVLHEEKCEKRIIDALE---LEPSI 362  
QY 172 AVLVASGGVANSFYIRALEILITNATOCTILCPPRLCTDNGIMIAMNGIEIRLPIAGIL 231  
DB 363 KEMVISGGVANSKYRLRLNINI VENKNULVCPPSLCTDNGVAMVMTGLBHRVG----- 418  
QY 232 HDIEGIRYBPCKP 244  
DB 419 -----RYDPPPP 425

RESULT 13  
AAG19285  
ID AAG19285 standard; protein; 444 AA.  
XX  
AC AAG19285;  
XX

DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21029.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
XX EPI033405-A2.  
PD 06-SBP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
PF  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136382P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139452P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147415P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149932P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.

```

PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153707P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161932P.
PR 29-OCT-1999; 99US-0162142P.

```

Query Match 28.8%; Score 398.5; DB 3; Length 444;  
 Best Local Similarity 38.7%; Pred. No. 1.4e-35;  
 Matches 98; Conservative 34; Mismatches 90; Indels 31; Gaps 6;

```

QY 1 MEAAHATIRL-NKKEPFLVLTISGGHCLALVGVSDPDLIGSLDIAPGMDLKYAR 59
DB 200 MEAAHATIRL-NKKEPFLVLTISGGHCLALVGVSDPDLIGSLDIAPGMDLKYAR 259
QY 60 RLSLIKRPECSTWSGKAIIEHLAKOGNRFHFDIKPPLHAKNCDFSFTGLQHTVDKIIMK 119
DB 260 WLGIDMH-----RSGCPAVEBELALEGDAKSVKFNPMKYHDCNFSYAGLKTQVRLAIEA 314
QY 120 KEKEBGEIKGQIILSSA-----ADIAATVOHTMACHLVKRTHRAILFCQKRDLLPQNN 171
DB 315 KE-----IDAKCPVSDATNEDRRNRADIAASFQRAVAVLHLEEKCEBALDIALB---LEPSI 367
QY 172 AVAVASGVASNFYIRALIELTNATQCTLLCPRLCTDNGIMIMANGIEELRAGLGL 231
DB 368 KHWIISGVASNKYVRLRLNINI VENKNLKLVCPPSLCTDNGVMVAMTGLEHFRVG----- 423
QY 232 HDIGIRPEKCP 244

```

DB 424 -----RYDPPP 430

#### RESULT 14

ID AAB31054 standard; protein; 251 AA.

AC AAB31054;

DT 24-FEB-2003 (first entry)

DE Ehrlichia ruminantium hypothetical o-sialoglycoprotein endopeptidase.

KW Vaccine; immunity; rickettsial infection; spotted fever; heart water;

KW typhus; pathogen; immunostimulant; antibacterial; lhwoif4; enzyme;

OS Ehrlichia ruminantium.

PN MO20026652-A2.

PD 29-AUG-2002.

PP 20-FEB-2002; 2002MO-US005772.

PR 20-FEB-2001; 2001US-0269944P.

PA (UYFL ) UNITV FLORIDA.

PI Barbet AF, Whitmire WW, Kamper SM, Simbi BH, Ganta RR;

PI Moreland AL, Mwangi DM, McGuire TC, Mahan SM;

XX MPI: 2002-723186/78.

DR N-PSDB; AAD48239.

PT New Ehrlichia ruminantium polynucleotides, useful as vaccines for

PS Claim 8; Page 67-68; 206pp; English.

CC The present invention relates to nucleic acid vaccines for conferring

CC immunity to rickettsial infection, including Ehrlichia ruminantium

CC (formerly Cowdria ruminantium). The invention also relates to novel B.

CC ruminantium polynucleotides and their corresponding proteins. Sequences

CC of the invention are useful for inducing immunity, particularly

CC protective immunity. They are also useful for detecting the presence of

CC B. ruminantium in a biological sample. They are useful in vaccines for

CC protecting animals or humans against rickettsial diseases, e.g. typhus,

CC spotted fever or heart water. Sequences of the invention are useful for

CC detecting antibodies to pathogens. The present sequence is B. ruminantium

CC hypothetical o-sialoglycoprotein endopeptidase (lhwoif4) protein

CC Sequence 251 AA;

Query Match 24.7%; Score 342; DB 5; Length 251;  
 Best Local Similarity 34.8%; Pred. No. 1.4e-29;  
 Matches 86; Conservative 39; Mismatches 90; Indels 32; Gaps 6;

```

QY 1 MEAAHATIRL-NKKEPFLVLTISGGHCLALVGVSDPDLIGSLDIAPGMDLKYAR 60
DB 20 LEAHTLILRMFDIDPFLVLTISGGHCLALVGVSDPDLIGSLDIAPGMDLKYAR 79
QY 61 LSLIKRPECSTWSGKAIIEHLAKOGNRFHFDIKPPLHAKNCDFSFTGLQHTVDKIIMK 120
DB 80 LNL-GVP-----GGPIIEKKSIMGDSKSFPLPRALINLGLCDFSPSGIKTAVRNIIVNQ 132
QY 121 EKEBGEIKGQIILSSAADIATVOHTMACHLVKRTHRAI-----LPKQKRDLLPQNNAVY 175
DB 133 KYIDN-----DFICNISASFQDCIGDILVIRITNAIMSAINKINK-----LV 177
QY 176 ASGVASNFYIRALIELTNATQCTLLCPRLCTDNGIMIMANGIEELRAGLGLHDI 235

```

Db 178 VTGVAANHLNRNISCVKDNFVLPTELCTDNGIMVGMAGIENLSKGY-----VS 232  
 Qy 236 GIRYEPK 242  
 Db 233 NLDVFPK 239

RESULT 15

ABU22934  
 ID ABU22934 standard; protein, 382 AA.

ABU22934;

19-JUN-2003 (first entry)

Protein encoded by Prokaryotic essential gene #8461.

Antisense; prokaryotic essential gene; cell proliferation; drug design.

Bordetella pertussis.

W0200277183-A2.

03-OCT-2002.

21-MAR-2002; 2002WO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,

Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

N-PSDB; ACA26804.

WPI; 2003-023926/02.

Claim 25; SEQ ID NO 50858; 1766bp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this CC patent did not form part of the printed specification, but was obtained CC in electronic format directly from WIPO at CC ftp.wipo.int/pub/published\_pct\_sequences

Sequence 382 AA;

Query Match 23.2%; Score 321.5; DB 6; Length 382;

Best Local Similarity 37.4%; Pred. 5.6e-27; Matches 88; Conservative 38; Mismatches 88; Indels 21; Gaps 8;

Qy 1 MEAHATTIRLTN-KYEFPPVLVLLISGHCILALVQGVSDPFLILGKSLDIAPGMDLKVAR 59

Db 148 LEGLHSLPLAARPRPFVALLVSGHQLVLDVGVREYELGELTDADAGAPFKSAK 207

Qy 60 RLSLIRHPECSTMSGKALEHLAKQGNRFHFDIKPPLHAKNCDSPFTGLQ-HYTDKITM 118

Db 208 LMGL-GYP-----GGPALARLAEOGDASRYDLFRPMLHSGDLDFSGIKTAVLTRV-- 258

Qy 119 KKEKEGIEKQIILSSADIAATVQHTMACHLVKRTHRAILFKORDLIPQNNAVLVASG 178

Db 259 KATRDGGELG--QDRADLAATQAAIVEVLAAKAIRL--KQYGL-----RLVAVG 308

Qy 179 GVAASFYIRALRILTNATQCTLLCPPRLCTDNGIMIMWNGIERLRAGIHLHD 233

Db 309 GVGANALRLAHLRALRKLRLARAYFPPLSLCTDNGMIFPAALRYKAGIADURE 363

Search completed: February 16, 2005, 13:05:56  
 Job time : 55.6902 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:43:41 ; Search time 15.962 Seconds  
(without alignments)  
1248.675 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHALTIRLTNKKVEPFLV.....DISKEVGASIKVPQLKMEI 267

Scoring table: BLASTSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	100.0	267	4 US-10-067-443-22	Sequence 22, Appl
2	1385	100.0	414	4 US-10-067-443-2	Sequence 2, Appl
3	1358.5	98.1	439	4 US-10-067-443-19	Sequence 19, Appl
4	409.5	29.6	463	4 US-10-067-443-3	Sequence 3, Appl
5	307	22.2	350	4 US-09-540-236-2726	Sequence 2726, Ap
6	291.5	21.0	401	4 US-09-252-991A-17312	Sequence 17312, A
7	280.5	20.3	357	4 US-09-543-681A-6513	Sequence 6513, Ap
8	279	20.1	342	1 US-08-087-797-3	Sequence 3, Appl
9	276	19.9	421	4 US-10-067-443-4	Sequence 4, Appl
10	276	19.9	421	4 US-10-067-443-28	Sequence 28, Appl
11	270.5	19.5	325	1 US-08-087-797-2	Sequence 2, Appl
12	268.5	19.4	343	4 US-09-489-039A-9221	Sequence 9221, Ap
13	261	18.8	363	4 US-09-107-532A-6609	Sequence 6609, Ap
14	254	18.3	336	3 US-08-961-121A-4	Sequence 4, Appl
15	250	18.1	335	3 US-08-961-083-52	Sequence 52, Appl
16	250	18.1	335	4 US-09-536-784-52	Sequence 52, Appl
17	248	17.9	336	3 US-09-066-512-2	Sequence 52, Appl
18	246	17.8	336	4 US-09-583-110-4857	Sequence 4857, Ap
19	243	17.5	327	4 US-10-067-443-5	Sequence 5, Appl
20	241	17.4	336	4 US-09-107-433-4221	Sequence 4221, Ap
21	236	17.0	273	4 US-09-710-279-728	Sequence 728, App
22	236	17.0	366	4 US-09-134-000C-4956	Sequence 4956, Ap
23	233	16.8	368	4 US-09-134-001C-3909	Sequence 3909, Ap
24	220	15.9	341	2 US-09-145-624-2	Sequence 2, Appl
25	218	15.7	344	4 US-09-198-452A-213	Sequence 213, Appl
26	218	15.7	360	4 US-09-438-185A-196	Sequence 196, App
27	211.5	15.3	344	4 US-09-602-777A-148	Sequence 148, App

28	194	14.0	340	4 US-10-067-443-6	Sequence 6, Appl
29	181.5	13.1	143	4 US-09-328-352-4387	Sequence 4387, Ap
30	169	12.2	292	4 US-09-724-623-81	Sequence 81, Appl
31	157	11.3	214	4 US-09-328-352-4609	Sequence 4609, Ap
32	94.5	6.8	480	4 US-09-583-110-5050	Sequence 5050, Ap
33	94.5	6.8	481	4 US-09-107-433-3197	Sequence 3197, Ap
34	83	6.0	42	4 US-09-902-540-13841	Sequence 13841, A
35	81.5	5.9	1087	1 US-08-264-002-5	Sequence 5, Appl
36	81	5.8	1072	4 US-09-902-540-15572	Sequence 15572, A
37	79	5.7	328	4 US-09-710-279-3008	Sequence 3008, Ap
38	79	5.7	336	3 US-09-134-001C-5549	Sequence 5549, Ap
39	79	5.7	445	3 US-08-083-945C-2	Sequence 2, Appl
40	79	5.7	445	3 US-08-083-945C-7	Sequence 7, Appl
41	79	5.7	578	4 US-09-252-991A-31318	Sequence 31318, A
42	79	5.7	1137	4 US-09-538-092-968	Sequence 968, App
43	78.5	5.7	503	4 US-09-252-991A-22790	Sequence 22790, A
44	78.5	5.7	1658	2 US-08-609-049A-13	Sequence 13, Appl
45	78.5	5.7	1658	3 US-09-170-996-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1					
US-10-067-443-22					
Sequence 22, Application US/10067443					
Patent No. 6642041					
GENERAL INFORMATION:					
APPLICANT: Bristol-Myers Squibb Company					
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN					
FILE OF INVENTION: SPINAL CORD, ME-1					
FILE REFERENCE: D0073 NP					
CURRENT APPLICATION NUMBER: US/10/067,443					
CURRENT FILING DATE: 2002-02-05					
PRIOR APPLICATION NUMBER: US 60/266,518					
PRIOR FILING DATE: 2001-02-05					
PRIOR APPLICATION NUMBER: US 60/282,814					
PRIOR FILING DATE: 2001-04-10					
NUMBER OF SEQ ID NOS: 71					
SOFTWARE: PatentIn version 3.0					
SEQ ID NO 22					
LENGTH: 267					
TYPE: PRT					
ORGANISM: homo sapiens					
US-10-067-443-22					
Query Match					
Best Local Similarity 100.0%; Score 1385; DB 4; Length 267;					
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MEAHALTIRLTNKKVEPFLVLLISGGHCLALVGVSPFLIGKSLDAPGMDLKVARR	60		
DB	1	MEAHALTIRLTNKKVEPFLVLLISGGHCLALVGVSPFLIGKSLDAPGMDLKVARR	60		
QY	61	LSLIRHPGCTMSGGKALIEHAKQGNRRFPDIKPLHAKNCDSPFTGQHTDTKIMKK	120		
DB	61	LSLIRHPGCTMSGGKALIEHAKQGNRRFPDIKPLHAKNCDSPFTGQHTDTKIMKK	120		
QY	121	EKEEGIEKGQILSSAADIAATVOHTMAGLVKTRTALIFCKQKRDLPQNNALVAASGV	180		
DB	121	EKEEGIEKGQILSSAADIAATVOHTMAGLVKTRTALIFCKQKRDLPQNNALVAASGV	180		
QY	181	ASNFIYRALBILTNATQCTLLCPPLRCTDNGIMIANNGIERLRAGLIGIRYE	240		
DB	181	ASNFIYRALBILTNATQCTLLCPPLRCTDNGIMIANNGIERLRAGLIGIRYE	240		
QY	241	PKCPGVDSIKVGEASIKVPQLKMEI	267		
DB	241	PKCPGVDSIKVGEASIKVPQLKMEI	267		
RESULT 2					
US-10-067-443-2					

Sequence 2, Application US/10067443  
Patent No. 6642041  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
LENGTH: 414  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-067-443-2

Query Match 100.0%; Score 1385; DB 4; Length 414;  
Best Local Similarity 100.0%; Pred. No. 7.3e-160;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAHLATRLTNKVEFPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 60  
DB 148 MEAHLATRLTNKVEFPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 207  
QY 61 LSLIKHPECSTMSGKAI EHLAKQGNRFPHDIKPLHAKNCDFSTGLQHTYTDKIIMKK 120  
DB 208 LSLIKHPECSTMSGKAI EHLAKQGNRFPHDIKPLHAKNCDFSTGLQHTYTDKIIMKK 267  
QY 121 EKEEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGGV 180  
DB 268 EKEEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGGV 327  
QY 181 ASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGIGILHDEIGIRY 240  
DB 328 ASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGIGILHDEIGIRY 387  
QY 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267  
DB 388 PKCPLGVDISKEVGEASIKVPOLKMEI 414

RESULT 3  
US-10-067-443-19  
Sequence 19, Application US/10067443  
Patent No. 6642041  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 19  
LENGTH: 439  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-067-443-19

Query Match 98.1%; Score 1358.5; DB 4; Length 439;  
Best Local Similarity 91.1%; Pred. No. 1.4e-156;  
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 MEAHLATRLTNKVEFPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 60  
DB 148 MEAHLATRLTNKVEFPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 207  
QY 61 LSLIKHPECSTMSGKAI EHLAKQGNRFPHDIKPLHAKNCDFSTGLQHTYTDKIIMKK 120  
DB 208 LSLIKHPECSTMSGKAI EHLAKQGNRFPHDIKPLHAKNCDFSTGLQHTYTDKIIMKK 267  
QY 121 EKEEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGGV 180  
DB 268 EKEEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGGV 327  
QY 181 ASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGIGILHDEIGIRY 240  
DB 328 ASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGIGILHDEIGIRY 387  
QY 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267  
DB 388 PKCPLGVDISKEVGEASIKVPOLKMEI 439

RESULT 4  
US-10-067-443-3  
Sequence 3, Application US/10067443  
Patent No. 6642041  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3  
LENGTH: 463  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-10-067-443-3

Query Match 29.6%; Score 409.5; DB 4; Length 463;  
Best Local Similarity 36.6%; Pred. No. 8.2e-41;  
Matches 100; Conservative 37; Mismatches 91; Indels 45; Gaps 6;

QY 1 MEAHLATRLTNKVEFPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 59  
DB 192 MEAHLATRLTNKVEFPFLVLLISGHCILALVQGVSDPFLLGKSLDIAPGMDLKVARR 251  
QY 60 LSLIKHPECSTMSGKAI EHLAKQGNRFPHDIKPLHAKNCDFSTGLQHTYTDKIIMKK 119  
DB 252 WLGUDMH-----RSGGPAVEBELALGDAKSVFENVPMKXHCNCSYAGLKTQYRALAIA 306  
QY 120 EKEEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCQKQDLPQNNAVLVASGGV 179  
DB 307 KE-----IRNRADIAASFQVAVVHLLEKCEERADMALE--LEPSIKHWISG 353  
QY 180 ASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGIGILHDEIGIRY 239  
DB 354 VASNFYIRALEILTNATQCTLLCPPRLCTDNGIMIANGIERLRAGIGILHDEIGIRY 403  
QY 240 E-----PKCPLGVDISKEVGEA 256  
DB 404 DEPPPATRPEVDYVYDLRPRWPLGEEVAKGRSRA 436

RESULT 5  
US-09-540-236-2726  
Sequence 2726, Application US/09540236  
Patent No. 6673910

```

; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540.236
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2726
; LENGTH: 350
; TYPE: PRT
; ORGANISM: M. catarrhalis
; US-09-540-236-2726

```

Query Match 22.2%; Score 307; DB 4; Length 350;

Best Local Similarity 33.3%; Pred. No. 1.7e-28; Matches 84; Conservative 37; Mismatches 91; Indels 40; Gaps 7;

```

QY 1 MEAAATLIRLT--NKVEPPLVLLISGGHCLLAVGVSDPFLLGKSLDIAPGMDLKV 58
DB 119 MEGHLLAPLASDPPSPFVCLVSGGHTLVADGVGVYQILGESIDDAVGECPDKTA 178
QY 59 RLSLIKHPECSTMSGKAIENHLAKQGNRFHDIKPLHAKNCDPSFTGLQHTVDKIM 118
DB 179 KMLGL-DYP-----GGPNIEKLANGNPHAYELPRPMQH-KGLDFSPGKMTALHNLIX 230
QY 119 KEKEBGEIKQILSSADIAATVQHTMACHLVKTRTHAILFCQKRDLLPQNNAVLVASG 178
DB 231 DTPNAQSDP-----ATRADIASFEYAVVDLVKKCTALQMTGIRQ-----LVVAG 277
QY 179 GVASNFYIRALBITLNAQCTLLCPPLCTDNGIMIAMNGIBRLRAG----- 227
DB 278 GVSANQTLRLITLTKQIDASVYAPTELTCTDNGAMTAYGFCRLSGQSDDLAVRCIP 337
QY 228 -----LGIHLD 233
DB 338 RMDWTMLGIEYD 349

```

RESULT 6

US-09-252-991A-17372

; Sequence 17372; Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 17372

; LENGTH: 401

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

; US-09-252-991A-17372

```

Query Match 21.0%; Score 291.5; DB 4; Length 401;
Best Local Similarity 35.4%; Pred. No. 1.6e-26;
Matches 87; Conservative 37; Mismatches 99; Indels 23; Gaps 9;

```

```

QY 120 KEKEBGEIKQILSSADIAATVQHTMACHLVKTRTHAILFCQKRDLLPQNNAVLVASG 179
DB 284 REVEAGDDEQ---TRCDIALAFQTAVENTLTKCRAL---KQTEL--KN---LVYAG 332
QY 180 VASNFYIRALBITLNAQCTLLCPPLCTDNGIMIAMNGIBRLRAGLIIHDIEGIRY 239
DB 333 VSAHQALRSGLKEMKEMGQVFYARPRPCTDNGAMTAYAGCQRLLAG---QHDGPAISV 389
QY 240 EPEKCP 245
DB 390 QPRWPM 395

```

RESULT 7

US-09-543-681A-6513

; Sequence 6513; Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543.681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 6513

; LENGTH: 357

; TYPE: PRT

; ORGANISM: Proteus mirabilis

; US-09-543-681A-6513

Query Match 20.3%; Score 280.5; DB 4; Length 357;

Best Local Similarity 32.7%; Pred. No. 3e-25;

Matches 81; Conservative 35; Mismatches 101; Indels 31; Gaps 8;

```

QY 1 MEAAATLIRLTNKV-EPPPLVLLISGGHCLLAVGVSDPFLLGKSLDIAPGMDLKV 59
DB 129 MEGHLLAPMLEBETKDPFPVALLVSGGHTLVADGVGVYQILGESIDDAVGECPDKTA 188
QY 60 RLSLIKHPECSTMSGKAIENHLAKQGNRFHDIKPLHAKNCDPSFTGLQHTVDKIM 119
DB 189 LIGL-DYP-----GGPLSKMAQGVBRFVFPFPMTRPGIDRSFGSLKTPAANTIQ 241
QY 120 KEKEBGEIKQILSSADIAATVQHTMACHLVKTRTHAILFCQKRDLLPQNNAVLVASG 178
DB 242 NDDSE-----QTRADIAFAFEDAVVDTLAIKCRRA-----LEQTFKRLVMAG 284
QY 179 GVASNFYIRALBITLNAQCTLLCPPLCTDNGIMIAMNGIBRLRAGL-GIHLIDIEI 237
DB 285 GVSANQTLRLITLTKQIDASVYAPTELTCTDNGAMTAYGFCRLSGQSDDLAVRCIP 340
QY 238 RYEPKCP 245
DB 341 TVRPRWPL 348

```

RESULT 8

US-08-087-797-3

; Sequence 3; Application US/08087797

; Patent No. 5543312

; GENERAL INFORMATION:

; APPLICANT: Mellors, Alan

; APPLICANT: Io, Regie Y.C.

; TITLE OF INVENTION: Pasteurella Haemolytica

; TITLE OF INVENTION: Glycoproteinase

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.

; STREET: 1211 East Morehead Street,



CITY: Charlotte  
 STATE: No. 5543312th Carolina  
 COUNTRY: United States  
 ZIP: 28234  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/087,797  
 FILING DATE: 14-JUL-1993  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Layton, Jr., Samuel G.  
 REGISTRATION NUMBER: 22807  
 REFERENCE/DOCKET NUMBER: 3374-80  
 TELEPHONE: 704 377 1561  
 TELEFAX: 704 334 2014  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 342 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-087-797-3

Query Match 20.1%; Score 279; DB 1; Length 342;  
 Best Local Similarity 33.6%; Pred. No. 4.2e-25;  
 Matches 83; Conservative 32; Mismatches 108; Indels 24; Gaps 8;

QY 1 MEAHALITRL-TNKVEPPLVLLISGGHCLALVGVSDPFLIGSLDIAFGMDLVAR 59  
 DB 112 MEHILALMLEDNPEPFVALVVGSGHTQLISVIGIYGLIGESIDDAAGSAFDTAK 171  
 QY 60 RLSLIKHPECSTMSGKAIENHAKGNRPHFDIKPPLHAKNCDPFTGLQHTDKI 119  
 DB 172 LLGL-DYP-----GGPLISKRAAGTGRFVPRPMTDRPGIDBSFSLKTFPAANTIRD 224  
 QY 120 KEKEGIEKQILISADIAATVQHTMACHLVKTTRAILFCQORDLPONN-AVLVAG 178  
 DB 225 NXXXXGXTHDD---TRADIAAFEDAVVDTLMIKKRA-----LDQTFKRLVAG 272  
 QY 179 GVASNFYIRALEILTNATQCTLLCPPLCTDNGIMIANWGIERLAGIGLHDIEGIR 238  
 DB 273 GVSANRTLRAXLAEMKKRGEVFYARPEFTDNGAMIAVAGVRFKA--GATADL-GVS 329  
 QY 239 YEPKCP 245  
 DB 330 VRPWPL 336

RESULT 9  
 US-10-067-443-4  
 Sequence 4, Application US/10067443  
 Patent No. 6642041  
 GENERAL INFORMATION:  
 APPLICANT: Bristol-Myers Squibb Company  
 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED  
 TITLE OF INVENTION: SPINAL CORD, MP-1  
 FILE REFERENCE: D0073 NP  
 CURRENT APPLICATION NUMBER: US/10/067,443  
 CURRENT FILING DATE: 2002-02-05  
 PRIOR APPLICATION NUMBER: US 60/266,518  
 PRIOR FILING DATE: 2001-02-05  
 PRIOR APPLICATION NUMBER: US 60/282,814  
 PRIOR FILING DATE: 2001-04-10  
 NUMBER OF SEQ ID NOS: 71  
 SOFTWARE: Patentin version 3.0  
 SEQ ID NO 4  
 LENGTH: 421

TYPE: PRT  
 ORGANISM: Caenorhabditis elegans  
 US-10-067-443-4

Query Match 19.9%; Score 276; DB 4; Length 421;  
 Best Local Similarity 30.4%; Pred. No. 1.4e-24;  
 Matches 85; Conservative 52; Mismatches 107; Indels 36; Gaps 8;

QY 1 MEAHALITRL-TNKVEPPLVLLISGGHCLALVGVSDPFLIGSLDIAFGMDLVAR 59  
 DB 134 MEHILALMLEDNPEPFVALVVGSGHTQLISVIGIYGLIGESIDDAAGSAFDTAK 171  
 QY 60 RLSLIKHPECSTMSGKAIENHAKGNRPHFDIKPPLHAKNCDPFTGLQHTDKI 116  
 DB 194 QUGDL-GSEFPGIHVAAVEILASRASADGHLRYPLFLNVPKANNPFOIKSYTLNLE 252  
 QY 117 IMKEKEGIEKQILISADIAATVQHTMACHLVKTTRAILFCQORDLPONN-AVLV 176  
 DB 253 RLKNSETSID-----IPDFCASIQNTVAHISSKLHIFPESLSEQETLPKQ---LVI 302  
 QY 177 GGVASNFYIRALEILTNATQCTLLCPPLCTDNGIMIANWGIERLAGIGLHDIEG 236  
 DB 303 GGVANAOYIFGALISLSAAHVVTTIKVLLSLCTDNAEMIAVSGL-----LMLVNRSEA 356  
 QY 237 IRYEP-----KCPLGVDISKVEGASIKVQLKM 265  
 DB 357 IMWRPNIDPTIYAHARSDIGTDASSET-----IDTPRRKL 392

RESULT 10  
 US-10-067-443-28  
 Sequence 28, Application US/10067443  
 Patent No. 6642041

QY 1 MEAHALITRL-TNKVEPPLVLLISGGHCLALVGVSDPFLIGSLDIAFGMDLVAR 59  
 DB 134 MEHILALMLEDNPEPFVALVVGSGHTQLISVIGIYGLIGESIDDAAGSAFDTAK 171  
 QY 60 RLSLIKHPECSTMSGKAIENHAKGNRPHFDIKPPLHAKNCDPFTGLQHTDKI 116  
 DB 194 QUGDL-GSEFPGIHVAAVEILASRASADGHLRYPLFLNVPKANNPFOIKSYTLNLE 252  
 QY 117 IMKEKEGIEKQILISADIAATVQHTMACHLVKTTRAILFCQORDLPONN-AVLV 176  
 DB 253 RLKNSETSID-----IPDFCASIQNTVAHISSKLHIFPESLSEQETLPKQ---LVI 302  
 QY 177 GGVASNFYIRALEILTNATQCTLLCPPLCTDNGIMIANWGIERLAGIGLHDIEG 236  
 DB 303 GGVANAOYIFGALISLSAAHVVTTIKVLLSLCTDNAEMIAVSGL-----LMLVNRSEA 356  
 QY 237 IRYEP-----KCPLGVDISKVEGASIKVQLKM 265

Query Match 19.9%; Score 276; DB 4; Length 421;  
 Best Local Similarity 30.4%; Pred. No. 1.4e-24;  
 Matches 85; Conservative 52; Mismatches 107; Indels 36; Gaps 8;

QY 1 MEAHALITRL-TNKVEPPLVLLISGGHCLALVGVSDPFLIGSLDIAFGMDLVAR 59  
 DB 134 MEHILALMLEDNPEPFVALVVGSGHTQLISVIGIYGLIGESIDDAAGSAFDTAK 171  
 QY 60 RLSLIKHPECSTMSGKAIENHAKGNRPHFDIKPPLHAKNCDPFTGLQHTDKI 116  
 DB 194 QUGDL-GSEFPGIHVAAVEILASRASADGHLRYPLFLNVPKANNPFOIKSYTLNLE 252  
 QY 117 IMKEKEGIEKQILISADIAATVQHTMACHLVKTTRAILFCQORDLPONN-AVLV 176  
 DB 253 RLKNSETSID-----IPDFCASIQNTVAHISSKLHIFPESLSEQETLPKQ---LVI 302  
 QY 177 GGVASNFYIRALEILTNATQCTLLCPPLCTDNGIMIANWGIERLAGIGLHDIEG 236  
 DB 303 GGVANAOYIFGALISLSAAHVVTTIKVLLSLCTDNAEMIAVSGL-----LMLVNRSEA 356  
 QY 237 IRYEP-----KCPLGVDISKVEGASIKVQLKM 265

Db 357 IMWRPNIDPITYAHARSDIGTADASSEI-----IDTPRRRL 392

## RESULT 11

US-08-087-797-2  
; Sequence 2, Application US/08087797  
; Patent No. 5543312

GENERAL INFORMATION:  
APPLICANT: Mellors, Alan

APPLICANT: Lo, Reggie Y.C.  
APPLICANT: Abdullah, Khalid M.

TITLE OF INVENTION: Pasteurella Haemolytica  
TITLE OF INVENTION: Glycoprotease

TITLE OF INVENTION: Gene and the Purified Enzyme  
NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.

STREET: 1211 East Morehead Street,  
CITY: Charlotte

STATE: No. 5543312th Carolina  
COUNTRY: United States

ZIP: 28234

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,797

FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Layton, Jr., Samuel G.

REGISTRATION NUMBER: 22807  
REFERENCE/DOCKET NUMBER: 3374-80

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 704 377 1561

TELEFAX: 704 334 2014  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 325 amino acids

TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-087-797-2

Query Match 19.5%; Score 270.5; DB 1; Length 325;  
Best Local Similarity 35.0%; Pred. No. 4,2e-24;

Matches 79; Conservative 26; Mismatches 102; Indels 19; Gaps 6;

QY 1 MEAHALTIRL-TNKVEPPLVLLISGHCLLALVGVSDPFLIGKSLDIAPGMDLKVAR 59  
DB 112 MEGHLAPMLEDNPPAPFFVALLVSGHTQLSVTIGQYELGESIDDAAGAFDKTK 171

QY 60 RLSLIKHPECSYMSGKAIHLAKOGRPHFDIKPPLHAKKCDFFSGTGLQHTVDKIMK 119  
DB 172 LLGL-DYF-----GGPMLSKWASQTBGRFVPRPMTDRPGLDFSGLTKTPAANTIK 224

QY 120 KEKEGIEKGQILSSADIAATVQHTMACHLVKTIRRALIFCKORDLPONNAVIVASGG 179  
DB 225 NLNENGELEDO--TKCDIAIAFOQAV-----VDITLIKCK-RALSQYGYKRLVMAGG 273

QY 180 VASNFYIRALIELTNATQCTLLCPPRLCTDNGIMIMANGIERLR 225  
DB 274 VSANKQIRADLAEMWKGLKGEVFPYRPFCTDNGMIAVGTGLRK 319

## RESULT 12

US-09-489-039A-9221  
; Sequence 9221, Application US/09489039A  
; Patent No. 6610835

GENERAL INFORMATION:

APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A

CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747

PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 9221  
LENGTH: 343

TYPE: PRN  
ORGANISM: Klebsiella pneumoniae

US-09-489-039A-9221

Query Match 19.4%; Score 268.5; DB 4; Length 343;  
Best Local Similarity 32.7%; Pred. No. 8,1e-24;

Matches 81; Conservative 31; Mismatches 105; Indels 31; Gaps 8;

QY 1 MEAHALTIRL-TNKVEPPLVLLISGHCLLALVGVSDPFLIGKSLDIAPGMDLKVAR 59  
DB 118 MEGHLAPMLEDNPPAPFFVALLVSGHTQLSVTIGQYELGESIDDAAGAFDKTK 177

QY 60 RLSLIKHPECSYMSGKAIHLAKOGRPHFDIKPPLHAKKCDFFSGTGLQHTVDKIMK 119  
DB 178 LLGL-DYF-----GGPMLSKWASQTBGRFVPRPMTDRPGLDFSGLTKTPAANTIK 230

QY 120 KEKEGIEKGQILSSADIAATVQHTMACHLVKTIRRALIFCKORDLPONN-AVLVASG 178  
DB 231 NGDDE-----QRRADIARAFEDAVVDTLMKCRRA-----LBQTSFKRLVMAG 273

QY 179 GVASNFYIRALIELTNATQCTLLCPPRLCTDNGIMIMANGIERLRAGIGLHIDE-GI 237  
DB 274 GVSANRTLRALKLAEMWKGRGEVFPYRPFCTDNGMIAVAGVRLQTA---KAEIGV 329

QY 238 RYEPKCP 245  
DB 330 TYRPRMPL 337

## RESULT 13

US-09-107-532A-6609  
; Sequence 6609, Application US/09107532A  
; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street  
CITY: Waltham

STATE: Massachusetts  
COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC  
OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571

FILING DATE: 3-Jul-97  
ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Denke  
REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 6609:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 363 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...363  
SEQUENCE DESCRIPTION: SEQ ID NO: 6609:  
US-09-107-512A-6609

Query Match 18.8%; Score 261; DB 4; Length 363;  
Best Local Similarity 33.8%; Pred. No. 7, 3e-23;  
Matches 75; Conservative 37; Mismatches 78; Indels 32; Gaps 9;

QY 1 MEAAHATIRLTNKKVEPPLVLLISGGCHLALVQGVSDFLLGKSLDAPGMDLVARR 60  
DB 140 MAGHIVARLVKPFQPLMALLVSGHTELVYMQEDSGYEIIIGETRDAAAGAYDKGRV 199  
QY 61 LSLIKHPECSTMSGKAIEHLAKOG-NRPHDIKPEPLHAKNCDPSTGLOHTDKIIMK 119  
DB 200 LGL-----SYPSKEIDQLAHQGRDNYH--PRAMIEDYDPSFSGLSKSAFINLVN 250  
QY 120 -KEKEGIEKGQILSSADIATVQHTWACHLVKTRTHAILFCOKRDLLPQNNAV--LVA 176  
DB 251 AQQGEGEDDKA-----DIAASFQASVYDVLTITLRA---C-----QYTPKQVLV 293  
QY 177 SGVAASFYIRRALBITNA--TQCTLLCPPELCTDNGIMI 216  
DB 294 AGVAANQGLREGLOALSAKLPVELVLPRLCGDNAMMI 335

RESULT 14  
US-08-987-121A-4  
Sequence 4, Application US/08987121A  
GENERAL INFORMATION:  
APPLICANT: Hoskins, Jo Ann  
APPLICANT: Tang, Joseph Chion-Chung  
TITLE OF INVENTION: Treadway, Patti Jean  
TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: U.S.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/987,121A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Webster, Thomas D  
REGISTRATION NUMBER: 39,872  
REFERENCE/DOCKET NUMBER: X-  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3334  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:

LENGTH: 336 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
US-08-987-121A-4

Query Match 18.3%; Score 254; DB 3; Length 336;  
Best Local Similarity 32.4%; Pred. No. 4, 6e-22;  
Matches 73; Conservative 36; Mismatches 86; Indels 30; Gaps 8;

QY 1 MEAAHATIRLTNKKVEPPLVLLISGGCHLALVQGVSDFLLGKSLDAPGMDLVARR 60  
DB 115 MAGHILMAQSVPEPLPPLALVSGHTELVYSBAGDYKIVETRDAAAGAYDKGRV 174  
QY 61 LSLIKHPECSTMSGKAIEHLAKOGNRPHDIKPEPLHAKNCDPSTGLOHTDKIIMK 120  
DB 175 MGL-----TYPAGEIDELAHQGHDI-YDPPRAVIKEDNIEFSGLSAPINLHNA 226  
QY 121 EKEGIEKGQILSSADIATVQHTWACHLVKTRTHAILFCOKRDLLPQNNAVLVAAGV 180  
DB 227 E-----QKESLST-EDLQASFQAAVMDILMAKTKAL-----EKYPVK--TLVYAGV 272  
QY 181 ASNFYIRRALBITNAVQCTLLCPPELCTDNGIMI-----WN 219  
DB 273 AANKGLRERL--ATBITDVNVIIIPRLCGDNAGMIAVASVSEWN 315

RESULT 15  
US-08-961-083-52  
Sequence 52, Application US/08961083  
Patent No. 6159469  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 335 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
US-08-961-083-52

Query Match 18.1%; Score 250; DB 3; Length 335;  
Best Local Similarity 32.0%; Pred. No. 1, 4e-21;  
Matches 72; Conservative 36; Mismatches 87; Indels 30; Gaps 8;

```

QY 1 MEAHATIRLTNKVEFPFLVLLISGGHCLALVQVSDPFLIGSGUDTAPGMDLKVARR 60
Db 114 MAGHLMMAQSVPEFPLALALVSGHTELYVSEAGDYKIVGETRDDAVGEAYDKGRV 173
QY 61 LSLIKHPECSTMSGKALEHLAKQGNRPHFDIKPPLHAKNCDFSFTGLQHVTDKIMCK 120
Db 174 MGL-----TYPAGREIDELAHQODI-YDEPRAMIKEDNLEFSPSGLKSAFINLHNHNA 225
QY 121 EKEGIEKGQILSSADIAATVQHTMACHLVKTTRAILFCQORDLLPQNNAVLVASGV 180
Db 226 E-----QKGSLSLST-BDLCASFOAAVMDILMAKTKKAL-----EKYPVK--ILVVAGGV 271
QY 181 ASNFYIRALEILMNAQCTLLCPPRLCTDNGIMIA-----WN 219
Db 272 AANKGLRERL--AAETIDVKVITIPPLRLCGDNAGMIAIASVSXWN 314

```

Search completed: February 16, 2005, 13:10:25  
 Job time : 16.962 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 13:08:47 ; Search time 154.686 Seconds  
(without alignments)  
563.995 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414

Perfect score: 1385  
Sequence: 1 MEAHLITRLTNKVEFPFLV.....DISKVGASIKVPQLKMEI 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	100.0	267	14	US-10-067-443-22 Sequence 22, Appl
2	1385	100.0	267	15	US-10-649-273-22 Sequence 22, Appl
3	1385	100.0	267	15	US-10-651-722-22 Sequence 22, Appl
4	1385	100.0	414	14	US-10-067-443-22 Sequence 2, Appl1
5	1385	100.0	414	15	US-10-649-273-22 Sequence 2, Appl1
6	1385	100.0	414	15	US-10-651-722-22 Sequence 2, Appl1
7	1358.5	98.1	439	14	US-10-067-443-19 Sequence 19, Appl
8	1358.5	98.1	439	15	US-10-649-273-19 Sequence 19, Appl
9	1358.5	98.1	439	15	US-10-651-722-19 Sequence 19, Appl
10	1348	97.2	414	14	US-10-012-140-5 Sequence 5, Appl1
11	1105	79.8	364	15	US-10-094-749-2039 Sequence 2039, Ap
12	422.5	30.5	445	15	US-10-424-599-209259 Sequence 209259, Ap
13	409.5	29.6	463	14	US-10-067-443-3 Sequence 3, Appl1

14	409.5	29.6	463	15	US-10-649-273-3	Sequence 3, Appl1
15	409.5	29.6	444	15	US-10-651-722-3	Sequence 3, Appl1
16	359	25.9	444	15	US-10-437-963-113732	Sequence 113732, A
17	342	24.7	251	14	US-10-081-051-9	Sequence 9, Appl1
18	321.5	23.2	382	14	US-10-282-122A-50858	Sequence 50858, A
19	317	22.9	333	14	US-10-012-140-25	Sequence 25, Appl1
20	314.5	22.7	312	15	US-10-282-122A-44499	Sequence 44499, A
21	308.5	22.3	348	15	US-10-282-122A-63156	Sequence 63156, A
22	304.5	22.0	343	15	US-10-282-122A-67227	Sequence 67227, A
23	294	21.2	347	14	US-10-012-140-24	Sequence 24, Appl
24	291.5	21.0	341	9	US-09-815-242-11798	Sequence 11798, A
25	291.5	21.0	342	15	US-10-282-122A-66200	Sequence 66200, A
26	289.5	20.9	342	9	US-09-815-242-11043	Sequence 11043, A
27	289.5	20.9	342	15	US-10-282-122A-58204	Sequence 58204, A
28	280.5	20.3	340	15	US-10-282-122A-68438	Sequence 68438, A
29	279.5	20.2	337	9	US-09-815-242-13780	Sequence 13780, A
30	279.5	20.2	337	15	US-10-282-122A-75455	Sequence 75455, A
31	279.5	20.2	337	15	US-10-282-122A-78161	Sequence 78161, A
32	277.5	20.0	335	15	US-10-282-122A-55404	Sequence 55404, A
33	277	20.0	309	15	US-10-282-122A-61538	Sequence 61538, A
34	277	20.0	341	15	US-10-282-122A-67993	Sequence 67993, A
35	276	19.9	421	14	US-10-067-443-4	Sequence 4, Appl1
36	276	19.9	421	14	US-10-067-443-28	Sequence 28, Appl1
37	276	19.9	421	15	US-10-649-273-4	Sequence 4, Appl1
38	276	19.9	421	15	US-10-649-273-28	Sequence 28, Appl1
39	276	19.9	421	15	US-10-651-722-4	Sequence 4, Appl1
40	276	19.9	421	15	US-10-651-722-28	Sequence 28, Appl1
41	275.5	19.9	337	9	US-09-815-242-10304	Sequence 10304, A
42	275.5	19.9	337	15	US-10-282-122A-56695	Sequence 56695, A
43	263.5	19.0	354	15	US-10-282-122A-65768	Sequence 65768, A
44	263	19.0	341	15	US-10-282-122A-69613	Sequence 69613, A
45	257	18.6	338	15	US-10-282-122A-57817	Sequence 57817, A

ALIGNMENTS

RESULT 1  
US-10-067-443-22  
; Sequence 22, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-067-443-22

Query Match 100.0%; Score 1385; DB 14; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.8e-139;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAHLITRLTNKVEFPFLVLLISGHCILALVQGVSPDLILGKSLIDIRGMLDKVAR 60  
DB 1 MEAHLITRLTNKVEFPFLVLLISGHCILALVQGVSPDLILGKSLIDIRGMLDKVAR 60  
QY 61 LSLIHPCSTMSGGKALHAKQNRFFHFKPLHAKKNDPFTGLQHTYTDKIIMKX 120  
DB 61 LSLIHPCSTMSGGKALHAKQNRFFHFKPLHAKKNDPFTGLQHTYTDKIIMKX 120  
QY 121 EKESGIEKGQILSSAADIATVQHTMACHLVKRTIRAILFCQRDILLPQNNAVLASGCV 180

Db 121 EKEGIEKQIISADIAATVQHTMACHLVYKTRHAILFCQKRDLLPNNNAVVASGCV 180  
QY 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLILHDIGIRYE 240  
Db 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLILHDIGIRYE 240  
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267  
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 2  
US-10-649-273-22  
; Sequence 22, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-649-273-22

Query Match 100.0%; Score 1385; DB 15; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.8e-139;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEAAHATIRLTNKKVEPFLVLLISGHCCLALVQGVSDFLILGKSIDIAPGMDLKVARR 60  
Db 1 MEAAHATIRLTNKKVEPFLVLLISGHCCLALVQGVSDFLILGKSIDIAPGMDLKVARR 60  
QY 61 LSLIKHPECSTWSGKAI EHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 120  
Db 61 LSLIKHPECSTWSGKAI EHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 120  
QY 121 EKEGIEKQIISADIAATVQHTMACHLVYKTRHAILFCQKRDLLPNNNAVVASGCV 180  
Db 121 EKEGIEKQIISADIAATVQHTMACHLVYKTRHAILFCQKRDLLPNNNAVVASGCV 180  
QY 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLILHDIGIRYE 240  
Db 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLILHDIGIRYE 240  
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267  
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 3  
US-10-651-722-22  
; Sequence 22, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05

; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-651-722-22

Query Match 100.0%; Score 1385; DB 15; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.8e-139;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEAAHATIRLTNKKVEPFLVLLISGHCCLALVQGVSDFLILGKSIDIAPGMDLKVARR 60  
Db 1 MEAAHATIRLTNKKVEPFLVLLISGHCCLALVQGVSDFLILGKSIDIAPGMDLKVARR 60  
QY 61 LSLIKHPECSTWSGKAI EHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 120  
Db 61 LSLIKHPECSTWSGKAI EHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 120  
QY 121 EKEGIEKQIISADIAATVQHTMACHLVYKTRHAILFCQKRDLLPNNNAVVASGCV 180  
Db 121 EKEGIEKQIISADIAATVQHTMACHLVYKTRHAILFCQKRDLLPNNNAVVASGCV 180  
QY 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLILHDIGIRYE 240  
Db 181 ASNFYIRALAILTNATQCTLLCPPRLCTDNGIMIANNGIERLRAGLILHDIGIRYE 240  
QY 241 PKCPLGVDISKVEGASIKVPOLKMEI 267  
Db 241 PKCPLGVDISKVEGASIKVPOLKMEI 267

RESULT 4  
US-10-067-443-2  
; Sequence 2, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-04-10  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-067-443-2

Query Match 100.0%; Score 1385; DB 14; Length 414;  
Best Local Similarity 100.0%; Pred. No. 3.4e-139;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MEAAHATIRLTNKKVEPFLVLLISGHCCLALVQGVSDFLILGKSIDIAPGMDLKVARR 60  
Db 148 MEAAHATIRLTNKKVEPFLVLLISGHCCLALVQGVSDFLILGKSIDIAPGMDLKVARR 207  
QY 61 LSLIKHPECSTWSGKAI EHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 120  
Db 208 LSLIKHPECSTWSGKAI EHLAKQGNRFHDIKPLHAKNCDFSFTGLQHTVDKTIIMKK 267  
QY 121 EKEGIEKQIISADIAATVQHTMACHLVYKTRHAILFCQKRDLLPNNNAVVASGCV 180

```
Db      268 EKEGIEKGQILSSAADIATVQHTMACHLVKRTTRAILFCRKORDLPPONNAVIVASGV 327
Qy      181 ASNFYIRRALBITLNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240
Db      328 ASNFYIRRALBITLNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 387
Qy      241 PKCPLGVDISKVEGASIKVPOLKMEI 267
Db      388 PKCPLGVDISKVEGASIKVPOLKMEI 414
```

```
RESULT 5
US-10-649-273-2
; Sequence 2, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2
```

```
Query Match      100.0%; Score 1385; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.4e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MEAHALTIRLTNKVEFPFLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLKVARR 60
Db      148 MEAHALTIRLTNKVEFPFLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLKVARR 207
Qy      61 LSLIHPEECSTWSGKAIENHLAKQGNRFHFDIKPPLHAKNCDPSTGLQHTYTDKIMKK 120
Db      208 LSLIHPEECSTWSGKAIENHLAKQGNRFHFDIKPPLHAKNCDPSTGLQHTYTDKIMKK 267
Qy      121 EKEGIEKGQILSSAADIATVQHTMACHLVKRTTRAILFCRKORDLPPONNAVIVASGV 180
Db      268 EKEGIEKGQILSSAADIATVQHTMACHLVKRTTRAILFCRKORDLPPONNAVIVASGV 327
Qy      181 ASNFYIRRALBITLNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240
Db      328 ASNFYIRRALBITLNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 387
Qy      241 PKCPLGVDISKVEGASIKVPOLKMEI 267
Db      388 PKCPLGVDISKVEGASIKVPOLKMEI 414
```

```
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2
```

```
Query Match      100.0%; Score 1385; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.4e-139;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MEAHALTIRLTNKVEFPFLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLKVARR 60
Db      148 MEAHALTIRLTNKVEFPFLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLKVARR 207
Qy      61 LSLIHPEECSTWSGKAIENHLAKQGNRFHFDIKPPLHAKNCDPSTGLQHTYTDKIMKK 120
Db      208 LSLIHPEECSTWSGKAIENHLAKQGNRFHFDIKPPLHAKNCDPSTGLQHTYTDKIMKK 267
Qy      121 EKEGIEKGQILSSAADIATVQHTMACHLVKRTTRAILFCRKORDLPPONNAVIVASGV 180
Db      268 EKEGIEKGQILSSAADIATVQHTMACHLVKRTTRAILFCRKORDLPPONNAVIVASGV 327
Qy      181 ASNFYIRRALBITLNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240
Db      328 ASNFYIRRALBITLNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 387
Qy      241 PKCPLGVDISKVEGASIKVPOLKMEI 267
Db      388 PKCPLGVDISKVEGASIKVPOLKMEI 414
```

```
RESULT 7
US-10-067-443-19
; Sequence 19, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-19
```

```
Query Match      98.1%; Score 1358.5; DB 14; Length 439;
Best Local Similarity 91.1%; Pred. No. 2.5e-136;
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;
```

```
Qy      1 MEAHALTIRLTNKVEFPFLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLKVARR 60
Db      148 MEAHALTIRLTNKVEFPFLVLLISGHCILALVQGVSDFLLGKSLDIAPGMDLKVARR 207
Qy      61 LSLIHPEECSTWSGKAIENHLAKQGNRFHFDIKPPLHAKNCDPSTGLQHTYTDKIMKK 120
Db      208 LSLIHPEECSTWSGKAIENHLAKQGNRFHFDIKPPLHAKNCDPSTGLQHTYTDKIMKK 267
Qy      121 EKEGIEKGQILSSAADIATVQHTMACHLVKRTTRAILFCRKORDLPPONNAVIVASGV 180
Db      268 EKEGIEKGQILSSAADIATVQHTMACHLVKRTTRAILFCRKORDLPPONNAVIVASGV 327
Qy      181 ASNFYIRRALBITLNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 240
Db      328 ASNFYIRRALBITLNATQCTLLCPPRCLCTDNGIMIANNGIERLRAGLILHIDIGIRYE 387
Qy      241 PKCPLGVDISKVEGASIKVPOLKMEI 267
Db      388 PKCPLGVDISKVEGASIKVPOLKMEI 414
```

Db 268 EKEGGIFLISKVEQINIPGLCKIAHFCRYEKGQILSSADIAATVQHTMACHLVKRTH 327  
Qy 156 RAIFCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPLCTDNGIM 215  
Db 328 RAIFCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPLCTDNGIM 387  
Qy 216 IAWNGIERLRAGLGIHLDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 267  
Db 388 IAWNGIERLRAGLGIHLDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 439

RESULT 8  
US-10-649-273-19  
; Sequence 19, Application US/10649273  
; Publication No. US20040043407A1

; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 19  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-649-273-19

Query Match 98.1%; Score 1358.5; DB 15; Length 439;

Best Local Similarity 91.1%; Pred. No. 2.5e-136;  
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

Qy 1 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDFLILGKSLDIAPGMDLKVARR 60  
Db 148 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDFLILGKSLDIAPGMDLKVARR 207  
Qy 61 LSLIHPECSTMSGGKAI EHLAKQGNRPHFDIKPPLHAKNCDSEFTGLQHTVDKTIIMKK 120  
Db 208 LSLIHPECSTMSGGKAI EHLAKQGNRPHFDIKPPLHAKNCDSEFTGLQHTVDKTIIMKK 267  
Qy 121 EKEGGI-----EKGQILSSADIAATVQHTMACHLVKRTH 155  
Db 268 EKEGGIFLISKVEQINIPGLCKIAHFCRYEKGQILSSADIAATVQHTMACHLVKRTH 327  
Qy 156 RAIFCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPLCTDNGIM 215  
Db 328 RAIFCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPLCTDNGIM 387  
Qy 216 IAWNGIERLRAGLGIHLDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 267  
Db 388 IAWNGIERLRAGLGIHLDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 439

RESULT 9  
US-10-651-722-19  
; Sequence 19, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05

; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 19  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-651-722-19

Query Match 98.1%; Score 1358.5; DB 15; Length 439;

Best Local Similarity 91.1%; Pred. No. 2.5e-136;  
Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

Qy 1 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDFLILGKSLDIAPGMDLKVARR 60  
Db 148 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDFLILGKSLDIAPGMDLKVARR 207  
Qy 61 LSLIHPECSTMSGGKAI EHLAKQGNRPHFDIKPPLHAKNCDSEFTGLQHTVDKTIIMKK 120  
Db 208 LSLIHPECSTMSGGKAI EHLAKQGNRPHFDIKPPLHAKNCDSEFTGLQHTVDKTIIMKK 267  
Qy 121 EKEGGI-----EKGQILSSADIAATVQHTMACHLVKRTH 155  
Db 268 EKEGGIFLISKVEQINIPGLCKIAHFCRYEKGQILSSADIAATVQHTMACHLVKRTH 327  
Qy 156 RAIFCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPLCTDNGIM 215  
Db 328 RAIFCKORDLLPQNNNAVVASGVASNFYIRRALEIITNATQCTLLCPPLCTDNGIM 387  
Qy 216 IAWNGIERLRAGLGIHLDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 267  
Db 388 IAWNGIERLRAGLGIHLDIEGIRYEPKCPPLGVDISKVEGASIKVPOLKMEI 439

RESULT 10

US-10-012-140-5  
; Sequence 5, Application US/10012140  
; Publication No. US20030009017A1

; GENERAL INFORMATION:  
; APPLICANT: Leiby, Kevin R.  
; APPLICANT: Kapeller-Libermann, Rosana  
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
; FILE REFERENCE: 381552004900  
; CURRENT APPLICATION NUMBER: US/10/012,140  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: 60/246,768  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,772  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,185  
; PRIOR FILING DATE: 2000-11-15  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-012-140-5

Query Match 97.3%; Score 1348; DB 14; Length 414;

Best Local Similarity 97.4%; Pred. No. 3.1e-135;  
Matches 260; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDFLILGKSLDIAPGMDLKVARR 60  
Db 148 MEAAHATIRLTNKKVEPFPVLVLLISGGHCLALVQGVSDFLILGKSLDIAPGMDLKVARR 207



QY	61	LSLHHPCSMMSGKALEHAKONRFFEDIKPPLAHKANDFSTGLQHVTDKIMMK	120
			:
Db	208	LSLHHPCSMMSGKALEHAKONRFFEDIKPPLAHKANDFSTGLQHVTDKIMMK	267
QY	121	EKEBIEKGQILLSSADIAATVQHTMACHLVYRTHRAILFCRKORDLLPONNAVLVASGV	180
	::		
Db	268	KQBSIEBKQILLSSADIAATVQHTMACHLVYRTHRAILFCRKORDLLPONNAVLVASGV	327
QY	181	ASNFIYRALLEILNATQCTLLCPPRLCTDNGIMIANNGIEBLRAGLGIIHDIIEGIRYE	240
Db	328	ASNFIYRALLEILNATQCTLLCPPRLCTDNGIMIANNGIEBLRAGLGIIHDIIEGIRYE	387
QY	241	PKCPGLGVDSKEVGASIKVPOLKKEI	267
Db	388	PKCPGLGVDSKEVGASIKVPOLKKEI	414

```

RESULT 11
US-10-094-749-2039
? Sequence 2039, Application US/10094749
? Publication No. US20030219741A1
? GENERAL INFORMATION:
? APPLICANT: ISOGAI, TAKAO
? APPLICANT: SUGIYAMA, TOMOYASU
? APPLICANT: OTSUKI, TETSUJI
? APPLICANT: WAKAMATSU, AI
? APPLICANT: SATO, HIROYUKI
? APPLICANT: ISHII, SHIZUKI
? APPLICANT: YAMAMOTO, JUN-ICHI
? APPLICANT: ISONO, YUDKO
? APPLICANT: HIO, YURI
? APPLICANT: OTSUKA, KAOBU
? APPLICANT: NAGAI, KEIICHI
? APPLICANT: IRE, RYOTARO
? APPLICANT: TAMECHIKA, ICHIRO
? APPLICANT: SEKI, NAOHICO
? APPLICANT: YOSHIKAWA, TETSUOMU
? APPLICANT: OTSUKA, MOTOKYUKI
? APPLICANT: NAGAHARI, KENJI
? APPLICANT: MASUHO, YASUHIKO
? TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
? FILE REFERENCE: 08435/0160
? CURRENT APPLICATION NUMBER: US/10/094,749
? CURRENT FILING DATE: 2002-03-12
? PRIOR APPLICATION NUMBER: 60/350,435
? PRIOR FILING DATE: 2002-01-24
? PRIOR APPLICATION NUMBER: JP 2001-328381
? PRIOR FILING DATE: 2001-09-14
? NUMBER OF SEQ ID NOS: 381
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 2039
? LENGTH: 364
? TYPE: PRT
? ORGANISM: Homo sapiens
US-10-094-749-2039

```

Query Match	79.8%	Score 1105	DB 15	Length 364
Best Local Similarity	98.6%	Prod. No. 2.6e-109		
Matches 214	Conservative	0	Mismatches 3	Indels 0
			Gaps	0
QY	1	MEAAHATIRLTNNKVEPPLVLLISGGHCLLALVGVSDPFLIGKSLIDIAPIGDMLDKVAAR	60	
Db	148	MEAAHATIRLTNNKVEPPLVLLISGGHCLLALVGVSDPFLIGKSLIDIAPIGDMLDKVAAR	207	
QY	61	LSLIRKHPCCSTMSGGKALIEHLAKQGNRRPFDIKPPLAHAKNCDPSTGQHYTDKIMKK	120	
Db	208	LPLIKHPCCSTMSGGKALIEHLAKQGNRRPFDIKPPLAHAKNCDPSTGQHYTDKIMKK	267	
QY	121	EKEEGIEKGQILSSADYIAATVOHMACTLVKTRTRAILFCRKORDLLPOMNAVAVASGGV	180	
Db	268	EKEEGIEKGQILSSADYIAATVOHMACTLVKTRTRAILFCRKORDLLPOMNAVAVASGGV	327	
QY	181	ASNFYIRRALEILTNAQTCTLCPPRLCTDNGIMIA	217	

Db 328 ASNFCIRALEILTNAQTCLLCPPLCTDNGIMIA 364

```

RESULT 12
US-10-424-599-209259
; Sequence 209259, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ_ID NOS: 285684
; SEQ_ID NO 209259
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_3098C.1.pep
; US-10-424-599-209259

```

Query Match	30.5%	Score 422.5;	DB 15;	Length 445;
Best Local Similarity	41.6%;	Pred. No. 4.1e-36;		
Matches 104;	Conservative 32;	Mismatches 89;	Indels 25;	Gaps 6;
Qy	1	MEAAATLIRLTNK-VEEPPVLVLISGGCHLTAIVQGVSDFFLLGKSLDIAPGDMIDYKAR	59	
		: : : : :       : : : : :       : : : : :       : : : : :       : : : : :		
Db	165	MEAAHAGRLLENDLQFFPMALLISGGHLLVLAARDGQYQLQSTTIDDAIGEAIVDKAK	224	
		: : : : :       : : : : :       : : : : :       : : : : :       : : : : :		
Qy	60	RLSLIKHBECSYMSGKAIEHLAKQGNRFHEDIKRPLHAKNCDPSFTGLQ-----HYTD	114	
		: : : : :       : : : : :       : : : : :       : : : : :       : : : : :		
Db	225	WLGGLDKR-----RSGPAIEKTLAMEGNAASVKSPIKMGKHDCNFSYAGLKYVRLAIES	279	
		: : : : :       : : : : :       : : : : :       : : : : :       : : : : :		
Qy	115	KIIMKEKEBEGIEKGQLLSPADIDATVAIVQHTMACHLYKTRHAILLFCQKORDLLPQNNAVL	174	
		: : : : :       : : : : :       : : : : :       : : : : :       : : : : :		
Db	280	KKIDAKIPISASNGDRL-SRADIAASFQRIAVLHLEERCEKRAIQWMLKMEPSIRH---L	335	
		: : : : :       : : : : :       : : : : :       : : : : :       : : : : :		
Qy	175	VASGGSVSNPIYRALIEILTNAOCTLLCPRPRLCTNGIMIAANGAERLPAIGILHDI	234	
		: : : : :       : : : : :       : : : : :       : : : : :       : : : : :		
Db	336	VWSGGVASNOYVVRARLDIMVVKQGLQVCEPPRLCTNGVIAVTAIGIEHFRMG-----	388	
		: : : : :       : : : : :       : : : : :       : : : : :       : : : : :		
Qy	235	EGIRYBEKCP	244	
		: : : : :       : : : : :       : : : : :       : : : : :       : : : : :		
Db	389	---RYDPPPP	395	

```

RESULT 13
US-10-067-443-3
; Sequence 3, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:10:32 ; Search time 4248.78 Seconds  
(without alignments)  
3045.000 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHALRIIRLNKVEPFLV.....DISKEVGASIKVPLKMEI 267

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 470823 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=x1h  
-Q=/cgsl1/USPTO.spool/US10649273/runat\_14022005\_114702\_16399/app\_query.fasta\_1.1429  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=humand4.cdi -LIST=45  
-LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US10649273.QCGN\_1\_1\_8655.qrunat\_14022005\_114702\_16399 -MCPU=6 -ICPU=3  
-NO\_MMAP -IARBOUDRY -NEG\_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*

1: gb Da:\*  
2: gb Hcg:\*  
3: gb In:\*  
4: gb Om:\*  
5: gb Ov:\*  
6: gb Pat:\*  
7: gb Ph:\*  
8: gb Pl:\*  
9: gb Pr:\*  
10: gb Ro:\*  
11: gb Sts:\*  
12: gb Sy:\*  
13: gb Un:\*  
14: gb Vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	100.0	1416	6	AR541929 Sequence
2	1385	100.0	1526	6	AR428809 Sequence
3	1385	100.0	1908	9	BC011904 Homo sapi
4	1385	100.0	2197	6	AR428803 Sequence

5	1358.5	98.1	1387	6	AR428808	AR428808 Sequence
6	1358.5	98.1	1387	9	HS295148	AJ295148 Homo sapi
7	1348	97.3	1245	6	AX664697	AX664697 Sequence
8	1348	97.3	1820	6	AX664695	AX664695 Sequence
9	1231	88.9	1844	10	BC058172	BC058172 Mus muscu
10	1224	88.4	1017	10	BC038910	BC038910 Mus muscu
11	1204	86.9	2208	6	AX713716	AX713716 Sequence
12	1204	86.9	2208	6	AX055441	AX055441 Homo sapi
13	1116	80.6	1546	10	BC078974	BC078974 Rattus no
14	950.5	68.6	1109	10	BC051211	BC051211 Mus muscu
15	939	67.8	1522	5	BX934991	BX934991 Gallus ga
16	938	67.7	860	5	BX930694	BX930694 Gallus ga
17	934	67.4	1558	5	BX930963	BX930963 Gallus ga
18	831.5	60.0	8415	6	AC013468	AC013468 Homo sapi
19	824.5	59.5	14364	6	AR428807	AR428807 Sequence
20	666	48.1	249601	2	AC114153	AC114153 Rattus no
21	666	48.1	308652	2	AC121478	AC121478 Rattus no
22	652.5	47.1	256751	10	AC122925	AC122925 Mus muscu
23	610	44.0	875	6	CO721898	CO721898 Sequence
24	488.5	35.3	1474	3	AK113378	AK113378 Clona int
25	477.5	34.5	121251	5	AL591593	AL591593 Zebrafish
26	435	31.4	117322	5	AL672217	AL672217 Zebrafish
27	415.5	30.0	1576	3	AY051882	AY051882 Drosophi
28	415.5	30.0	1601	6	CO606432	CO606432 Sequence
29	415.5	30.0	3656	6	CO606431	CO606431 Sequence
30	415.5	30.0	14679	2	AC018262	AC018262 Drosophi
31	415.5	30.0	180263	3	AC010671	AC010671 Drosophi
32	415.5	30.0	207432	3	AE003513	AE003513 Drosophi
33	408	29.5	1672	8	AK070912	AK070912 Oryza sat
34	407.5	29.4	1443	8	AY024338	AY024338 Arabidops
35	407.5	29.4	1474	8	AY117283	AY117283 Arabidops
36	407.5	29.4	1567	8	AY063864	AY063864 Arabidops
37	401.5	29.0	1557	8	AY084577	AY084577 Arabidops
38	389	28.1	110000	1	AP006501	AP006501 Continuation (9 of
39	369	26.6	260271	1	AE017258	AE017258 Wolbachia
40	356.5	25.7	333800	1	SME591792	AL591792 Sinorhizo
41	356	25.7	20021	10	AF36796781	AF367967 Mus muscu
42	356	25.7	179252	10	AF131205	AF131205 Mus muscu
43	352	25.4	349116	1	AP003003	AP003003 Mesozitho
44	349	25.2	110000	1	AE017197_01	Continuation (2 of
45	342	24.7	756	6	AX685201	AX685201 Sequence

## ALIGNMENTS

RESULT 1	AR541929	Sequence 177 from patent US 6743619.	1416 bp	DNA	linear	PAT 08-OCT-2004
LOCUS	AR541929					
DEFINITION	Sequence 177 from patent US 6743619.					
ACCESSION	AR541929					
VERSION	AR541929.1	GI:53934009				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 1416)					
AUTHORS	Tang, Y. T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F., Zhang, J., Zhao, Q. A., Yang, Y., Xue, A. J., Wehrman, T., Wang, J. -R., Wang, D. and Drmanac, R. T.					
TITLE	Nucleic acids and polypeptides					
JOURNAL	Patent: US 6743619-A 177 01-JUN-2004;					
FEATURES	Location/Qualifiers					
source	1..1416					
ORIGIN	/organism="unknown"					
	/mol_type="genomic DNA"					

Alignment Scores:  
Pred. No.: 1,37e-126  
Score: 1385.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%

Length: 1416  
Matches: 267  
Conservative: 0  
Mismatch: 0  
Indels: 0

DB:	6	Gaps:	0
US-10-649-273-2_COPY_148_414 (1-267) x AR541929 (1-1416)			
Qy	1	MetGluAlaHisSalaleuThrTlIeaRgLeuThrAsnuLySValGluPheProPheLeuVal	20
Db	502	ATGAGAGGCTCATGCACCTTACTATTNGTGTGACCAATTAAAGTAAAGTATTCCTTTTTCAGTT	561
Qy	21	LeuLeuIleSergIyGIyHIScysLeuLeuAlaleuValGingIyValIserAspPheLeu	40
Db	562	CTTTTGAATTTCTGGAGGCTCACTGTCGTGTGGCATTAGTTCAGAGGAGTTTCAGATTTTCTG	621
Qy	41	LeuLeuGIyLySserLeuAspIleAlaIarProGIyAspMetLeuAspLySValAlaArg	60
Db	622	CTTCTTGGAAAGTCTTTGGACATAGCACACAGCTGACACGCTTGACCAAGTGGCAAGAGA	681
Qy	61	LeuSerLeuIleLySHIScysProGIyCysSerThrMetSerGIyLySValAlaIleGIyHIS	80
Db	682	CTTCTCTTTAAATTAACATCCAGAGTGTCTCCACCATGATGTGTGGAAAGCATTAGAACAT	741
Qy	81	LeuAlaLySgingIyAsnArgPheHisAspAspIleLySProProLeuHISHisAlaLyS	100
Db	742	TTGGCCAAACAAAGGAATATGATTTTCATTTTGACATCAAAACCTCCCTTGACATGCTTAA	801
Qy	101	AsnCySAspPheSerPheThrGIyLeuGlnHISValThrAspLySleIleMetLyS	120
Db	802	AATGTGATTTTTCTTTTACTGAGCTTCACACCTTACTGTGATTAATATGAAAAAG	861
Qy	121	GluLySgingIyGIyIleGluLySGIyGlnIleLeuSerSerAlaAlaAspIleAlaAla	140
Db	862	GAAAAAGGAGAGGATTTGAGAAAGGGCAATCTGCTTCCAGACAGACATTTGCTGCC	921
Qy	141	ThrValGlnHISThrMetAlcYSerHISLeuValLySArgThrHISArgAlaIleLeuPhe	160
Db	922	ACAGTACAGCACACATGCGATGTCATTTGTGAAAAAACAACATCGGCGTATTCGTGTT	981
Qy	161	CysLySglnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGIyGIyVal	180
Db	982	TGTATGACGAGAGACTGTGTACTCTCAAAATATATGACATGCTGTGCTGTGGTGTCTC	1041
Qy	181	AlaSerAsnPheTyrtIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr	200
Db	1042	GCAAGTAACTTCTATATCCGACAGCTGTGAAATTTTAAACAAAGCAACAGAGCACT	1101
Qy	201	LeuLeuCySProProProArgLeuCySThrAspAsnGIyIleMetIleAlaTrpAsnGIy	220
Db	1102	TTGTGTGTCTCTCTCCCAAGCATAGCATATATATGCACTTATTCATTCATGCAATGCT	1161
Qy	221	IleGluArgLeuArgAlaGIyLeuGIyIleLeuHISAspIleGIyGIyIleArgTyrtGlu	240
Db	1162	ATTGAAAGACATACGCTGCTGGCATTTTATCATGACATAGAAAGCATCCGCTATGAA	1221
Qy	241	ProLySAspProLeuGIyValAspIleSerLySGIyValGIyGlnAlaSerTleLyVal	260
Db	1222	CCAAAGTCTCTCTTGAGTACATATCAAAAGAGTGGAGAAAGCTTCCATTAAGTA	1281
Qy	261	ProGlnLeuLySmetGluIle 267	
Db	1282	CCACATTAATAAATGAGATA 1302	
RESULT 2			
LOCUS	AR428809	1526 bp	DNA linear PAT 18-DEC-2003
DEFINITION	Sequence 23 from patent US 6642041.		
ACCESSION	AR428809		
VERSION	AR428809.1	GI:40186595	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1526)		
AUTHORS	Chen,J., Feder,J.N., Nelson,T.C., Kyrstek,S.R. and Duclos,F.		
TITLE	Polynucleotides encoding a novel metalloprotease, MP-1		

JOURNAL	Patent: US 6642041-A 23 04-NOV-2003;
FEATURES	Location/Qualifiers
Source	1..1526
	/organism="unknown"
	/mol_type="genomic DNA"
ORIGIN	
Alignment Scores:	
Pred. No.:	1..5e-126
Score:	1385..00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	6
	Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x AR428809 (1-1526)	
QY	1 MetGuaIaHiaIaIeUthrIleAryLeuThraaUyVaIaGluPhropheUeUaI 20
Db	1 ATGGAGGGCTCAAGCACTTAATAGTGTGACCAATAAGTGAATTTCTTTTAAATT 60
QY	21 LeuLeuIleSerGIyGIyHiaCYsLeuLeuAlaLeuVaIGInGIyAlSerAaphLeu 40
Db	61 CTTTGGATTTCTGGAGGTCACCTGTCTGTGGCATTAAGTCAAGGAATTTCAATTTCTG 120
QY	41 LeuLeuGIyLySerLeuAspIleAlaProGIyAspMetLeuAspLyVaIAlaArg 60
Db	121 CTTTCTGGAAAGCTTTTGGACATAGCACACAGGTGCATCTTGGACAAAGGTGCMAAGA 180
QY	61 LeuSerLeuIleLyHiaProGIyCySerThrMetSerGIyGIyAlaIleGluHia 80
Db	181 CTTTCTTTAAATTAACATTCACAGAGTCCTCACCATGATGTGTGGAAAGCCATAGAAT 240
QY	81 LeuAlaLySGInGIyAsnArgPheHiaPheAspIleLyProProLeuHiaIaLyS 100
Db	241 TTGGCCCAACAAAGAAATAGATTTCATTTTGACATCAACCTCCCTTGACATCTAA 300
QY	101 AaCYsAspPheSerPheThrGIyLeuGInHiaSvaIThrAspLySIIeMetLyS 120
Db	301 AATTGTGATTTTCTTTTAAGTGAATTCGAACGTTACTGATGAATTAATGAAGAAAG 360
QY	121 GIuLySGInGIyGIyIleGluLySGInIleLeuSerSerAlaIaAspIleAla 140
Db	361 GAAAAAGAGGAAGGATTAAGAAAGGGCAAATCTGTCTTCAGCAGCATTTGCTGCC 420
QY	141 ThrValIGInHiaSThMetAlaCYsHiaSleuValLySArgThHiaArgAlaIleLeuPhe 160
Db	421 ACGATACGACACAAATGCGATGTCAATCTTGTGAAAGAACACATGGGCTAATTCGTTT 480
QY	161 CySLySGInArgAspLeuLeuProGInaAsnAlaValLeuVaIaIaSerGIyGIyAla 180
Db	481 TGTAAAGCAGAGAGACTTGTTCCTCAAAATTAATGACGATCTGGTTCATCTGTGGTGC 540
QY	181 AlaSerAaPheThyIleArgArgAlaLeuGIyIleLeuThraaAlaHiaArgInCYsThr 200
Db	541 GCAAGTAACCTTAAATATCGCAGAGCTCTGGAATTTTAAACAAACCAACACAGTGCATC 600
QY	201 LeuLeuCySProProProArgLeuCySThraaAsnGIyIleMetIleAlaTPraaNGly 220
Db	601 TTGTGTGTCTCTCTCCACAGACTATGCACTGATTAATGGCATTTATGATGCAATGAATG 660
QY	221 ILeGIuArgLeuArgAlaGIyLeuGIyIleLeuHiaSAspIleGIyIleArgYrGIu 240
Db	661 ATTTGAAAGACTACGAGTGTGGCTTTGGCATTTTAACATGACATGAAGGCAATCGCTATGAA 720
QY	241 ProLySCySProLeuGIyAlaAspIleSerLySGInVaIGIyAlaSerIleLyVaI 260
Db	721 CCAAAATGTCTCTTGGAGTAGACATATCAAAAAGAGTGTGAAGACCTTCATTAAGAATA 780
QY	261 ProGInLeuLyMetGIyIle 267
Db	781 CCACAAATTAATAATGAGATA 801

RESULT 3  
LOCUS BC011904  
DEFINITION Homo sapiens O-6-acylglycoprotein endopeptidase-like 1, mRNA (cDNA  
clone MGC:20293 IMAGE:4121450), complete cds.  
ACCESSION BC011904  
VERSION BC011904.2  
KEYWORDS GI:40225818  
SOURCE MGC.  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 1908)  
REFERENCE  
AUTHORS Klausner R.D., Collins P.S., Wagner L.D., Derge J.G.,  
Krausberg R.L., Feingold E.A., Grouse L.H., Schemm C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Blac N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L.,  
Scheetz T.E., Brownstein M.J., Ueding T.B., Toshiyuki S.,  
Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J.,  
Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J.,  
McKernan K.J., Malek J.A., Gunaratne P.H., Richards S.,  
Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Hellon E., Kettleman M., Madan A., Rodriguez S.,  
Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y.,  
Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D.,  
Dickson M.C., Rodriguez A.C., Grimwood J., Schmitt J., Myers R.M.,  
Butterfield Y.S., Krzywinski M.I., Skalska U., Small J.D.,  
Schnerch A., Schein J.E., Jones S.J. and Marra M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 1908)  
Strausberg R.  
Direct Submission  
Submitted (30-JUL-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NHI-MGC Project URL: <http://mgc.nci.nih.gov>  
On Dec 19, 2003 this sequence version replaced gi:15080281.  
REMARK  
COMMENT Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
DNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Feature  
Location/Qualifiers  
1. 1908  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:20293 IMAGE:4121450"  
/tissue\_type="Muscle, rhadomyosarcoma"

US-10-649-273-2\_copy\_148\_414 (1-267) x BC011904 (1-1908)  
ORIGIN  
misc\_feature  
215..1111  
/gene="OSGEP1"  
/note="Peptidase M2; Region: Glycoprotease family"  
/db\_xref="CD:pfam0814"  
Alignment Scores:  
Pred. No.: 1,97e-126 Length: 1908  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
/clone\_1ib="NHI MGC 17"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
1..1908  
/gene="OSGEP1"  
/db\_xref="LocustID:64172"  
104..1348  
/gene="OSGEP1"  
/codon\_start=1  
/product="OSGEP1 protein"  
/protein\_id="AAH1904.1"  
/db\_xref="GI:15080282"  
/db\_xref="LocustID:64172"  
/translation="MLIRKTVGPEKPKRYEPLRSPNPHGTLPLKXIVLGIET  
SCDDPRAAYVDENVNTGEAHISQTPVHAKTGIVPPAAQOQHRNIOIYQALAS  
GSPDSLSAIVATTIKGLALSLGVGSFSLQVLGQKKPPIPIHHEAHALTRLNK  
VPEPFLVLLISGHCLALVQGVSDPLGKSLIDIPGMLDVARSLIKRPEST  
MISGKRIEHLAKQGNFHDIPKPLHAKCDFSPYGLQVTDTKIMKEKEGIEG  
QLSSAADIAATVQHTMACHLVTRHRAILFCRKQDLDPNNALVLSGVASNPYIR  
RLLEITNATQCTLICPPRLCTDNGIMAMNCIERLRAGLGIHIDIGIRYKCPPL  
GVDISKEVGRASIKYPLKMEY"  
US-10-649-273-2\_copy\_148\_414 (1-267) x BC011904 (1-1908)  
ORIGIN  
misc\_feature  
215..1111  
/gene="OSGEP1"  
/note="Peptidase M2; Region: Glycoprotease family"  
/db\_xref="CD:pfam0814"  
Alignment Scores:  
Pred. No.: 1,97e-126 Length: 1908  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
/clone\_1ib="NHI MGC 17"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
1..1908  
/gene="OSGEP1"  
/db\_xref="LocustID:64172"  
104..1348  
/gene="OSGEP1"  
/codon\_start=1  
/product="OSGEP1 protein"  
/protein\_id="AAH1904.1"  
/db\_xref="GI:15080282"  
/db\_xref="LocustID:64172"  
/translation="MLIRKTVGPEKPKRYEPLRSPNPHGTLPLKXIVLGIET  
SCDDPRAAYVDENVNTGEAHISQTPVHAKTGIVPPAAQOQHRNIOIYQALAS  
GSPDSLSAIVATTIKGLALSLGVGSFSLQVLGQKKPPIPIHHEAHALTRLNK  
VPEPFLVLLISGHCLALVQGVSDPLGKSLIDIPGMLDVARSLIKRPEST  
MISGKRIEHLAKQGNFHDIPKPLHAKCDFSPYGLQVTDTKIMKEKEGIEG  
QLSSAADIAATVQHTMACHLVTRHRAILFCRKQDLDPNNALVLSGVASNPYIR  
RLLEITNATQCTLICPPRLCTDNGIMAMNCIERLRAGLGIHIDIGIRYKCPPL  
GVDISKEVGRASIKYPLKMEY"

QY 181 AlaseaenpheyTlleargargAlaleugluilleuThraAsnAlaThrgInCysThr 200  
DB 1085 GCAAGTACTCTTATATCCGAGAGCTCTGAAATTTTACAAAGCAACAGAGTGCAT 1144  
QY 201 LeuleuCySPProProProargleuCyThraAspanglylleMetilleaIatrapangly 220  
DB 1145 TTGTTGTGTCCTCCCTCCAGACTAGCATGATATGACATTAATGACATGAGAAATGAT 1204  
QY 221 lIegluargleuArgAlaelyleuglylleuHiaspilleuglylleargtyrglu 240  
DB 1205 ATTGAAGACACGAGCTGCTGGCATTTTACATGACATGAAGAGGACATCGCATATGA 1264  
QY 241 ProlyCySPProleuglyValAspilleserlysgluValglylualaserilleysVal 260  
DB 1265 CCAAAATGCTCTTGGAGTACATATCAAAAGAAAGTTGGAGAACTTCCATTAAGTA 1324  
QY 261 ProgluLeuLyMetgluile 267  
DB 1325 CCACAATTAATAATGAGATA 1345  
RESULT 4  
LOCUS AR428803 2197 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 1 from patent US 6642041.  
ACCESSION AR428803  
VERSION AR428803.1 GI:40188589  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2197)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, Mp-1  
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;  
FEATURES  
source 1..2197  
location/Qualifiers  
/organism="Unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,33e-126 Length: 2197  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x AR428803 (1-2197)  
QY 1 MetglualAHlAlaleuThrlleargleuThraAsnLyValglunpheProphleuVal 20  
DB 672 ATGAGGCTCATGACACTTACTATTAGTTGACCAATTAAGTAATTCCTTTTACTT 731  
QY 21 LeuleuilleserlyglyHiasCysleuLeuAlaleuValglylValaseraspheleu 40  
DB 732 CTTTGTATTTCTGGAGGTCACTGTCTGTGGCATTAAGTTCAGAGTTTCAATTTTCTG 791  
QY 41 LeuleuglylySerleuAspilleaIaproglyAspMetleuAspLyValaIaargarg 60  
DB 792 CTTCTTGGAAGTCTTTGGACATAGCACAGGTGACATCTTGAACAAGTTGGCAAGAA 851  
QY 61 LeuSerleuIlelyHiasProgluCySerThrMetSerlyglylyValaIaIlegluHias 80  
DB 852 CTTTCTTAATTAATAACATCCAGAGTCTCCACATGAGTGTGGAGAAAGCCATAGAACAT 911  
QY 81 LeuAlaLySGInglyAsnArgpheHiaspheAspilleysProProleuHiasIleValys 100  
DB 912 TTGGCCAAACAGAAATGATTTGATTTGACATCAACCTCCCTTGATCATGCTTAA 971  
QY 101 AsnCyAspPheSerPheThrglyleuGlnHiasValThraSplylleIleMetLylys 120  
DB 972 AATGTGATTTTCTTTTACTGACTTCAACGCTTACTGATTAATTAATGAAAAAG 1031

QY 121 GluLySGIngluGlylleglunpysglylnilleuSerSerAlaIaAspilleaIaIa 140  
DB 1032 GAAAAAGAGAGGATTTTGAAGAGGCAATCTCTCTTACAGCAGACATGCTGCTCC 1091  
QY 141 ThrValGlnHiasThrMetAlaCyshisIleuVallyAsrgThrhiasrgAlaIleleuphe 160  
DB 1092 ACAGTACAGCACACATAGCATGTGATCTTGTGAAAAGAACATCGGGCTATTCTGTTT 1151  
QY 161 CysLySGInaArgAspIleuLeuProgluAsnAlaIaIleuValaIaserlyglyVal 180  
DB 1152 TGTAAAGCAGAGACTGTTTACTCAAAATTAATGACGTGCTGTCATCTGTGTGTGTC 1211  
QY 181 AlaseaenpheyTlleargargAlaleugluilleuThraAsnAlaThrgInCysThr 200  
DB 1212 GCAAGTACTCTTATATCCGAGAGCTCTGGAATTTTACAAAGCAACAGAGTGCAT 1271  
QY 201 LeuleuCySPProProProargleuCyThraAspanglylleMetilleaIatrapangly 220  
DB 1272 TTGTTGTGTCCTCCCTCCAGACTAGCATGATTAATGACATTAATGATGATGGAATGAT 1331  
QY 221 lIegluargleuArgAlaelyleuglylleuHiaspilleuglylleargtyrglu 240  
DB 1332 ATTGAAGACACGAGCTGCTGGCATTTTACATGACATGAAGAGCATCCGCTATGA 1391  
QY 241 ProlyCySPProleuglyValAspilleserlysgluValglylualaserilleysVal 260  
DB 1392 CCAAAATGCTCTTGGAGTACATATCAAAAGAAAGTTGGAGAACTTCCATTAAGTA 1451  
QY 261 ProgluLeuLyMetgluile 267  
DB 1452 CCACAATTAATAATGAGATA 1472  
RESULT 5  
LOCUS AR428808 1387 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 21 from patent US 6642041.  
ACCESSION AR428808  
VERSION AR428808.1 GI:40188594  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1387)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, Mp-1  
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;  
FEATURES  
source 1..1387  
location/Qualifiers  
/organism="Unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 5,43e-124 Length: 1387  
Score: 1358.50 Matches: 266  
Percent Similarity: 91.10% Conservative: 0  
Best Local Similarity: 91.10% Mismatches: 1  
Query Match: 98.09% Indels: 25  
Gaps: 1  
US-10-649-273-2\_COPY\_148\_414 (1-267) x AR428808 (1-1387)  
QY 1 MetglualAHlAlaleuThrlleargleuThraAsnLyValglunpheProphleuVal 20  
DB 465 ATGAGGCTCATGACACTTACTATTAGTTGACCAATTAAGTAATTCCTTTTACTT 524  
QY 21 LeuleuilleserlyglyHiasCysleuLeuAlaleuValglylValaseraspheleu 40  
DB 525 CTTTGTATTTCTGGAGGTCACTGTCTGTGGCATTAAGTTCAGAGTTTCTG 584  
QY 41 LeuleuglylySerleuAspilleaIaproglyAspMetleuAspLyValaIaargarg 60

```

Db      585 CTTCTTGGAAAAGTCTTTGACATAGCACACAGATGATCATGCTTGAACAAGTGGCAAGAAGA 644
Qy      61 LeuSerLeuIleIysHisProGluCysSerThrMetSerGlyIleValAlaIleGluHis 80
Db      645 CTTTCTTTAATAAACAATCCAGAGTGTCTCCACCATGAGTGGGGAAGCCATAGACAT 704
Qy      81 LeuAlaIysGlnGlyAsnArgPheHisPheAspIleIysProProLeuHisHisAlaIys 100
Db      705 TTGGCCAAACAAGAAATAGATTTCATTTCATTCATCAACCTCCCTTCATCATCTAA 764
Qy      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db      765 AATTGATATTTTCTTTACTGACATTCACACGTTACTGATAAATAATATAAGAAAG 824
Qy      121 GluIysGluGluGlyIle----- 126
Db      825 GAAAAAAGAGAGAGATATTTCTAATTAAGTAAAGTTGAACAGATAATATTCCTGATTTG 884
Qy      127 -----GluIysGlyGlnIleLeuSerSerAla 135
Db      885 TGCCATAAATAAGTCTGCTCATTTCTGCAGGTATGAGAAAGGGCAATCTCTTCACACA 944
Qy      136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155
Db      945 GCAGACATTTGCTGCACAGTACACACACAAATGGCATGTCATCTTGAAAAAACAACAT 1004
Qy      156 ArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175
Db      1005 CGGGCTATTTCTGTTTGTAAAGCAGAGACTGTGTTACTCTCAAAAATAATGACGTACTGTT 1064
Qy      176 AlaSerGlyIysValAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsn 195
Db      1065 GCATCTGTGTGTGTGCGAAGTAACTCTTAATCCGACAGCTCTGGAATTTTAAACAAC 1124
Qy      196 AlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMet 215
Db      1125 GCAACACAGTGCATTTGTGTGTCTCTCTCCAGACATATGACATGATATGAGCATTTTG 1184
Qy      216 IleAlaIleThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235
Db      1185 ATTGATGAGATGATGATTAAGAAAGACTAGTGTGGCTGGGCAATTTTACATGACATAGAA 1244
Qy      236 GlyIleLeuArgIleGluProLysCysProLeuGlyValAspIleSerLysGluValGlyLys 255
Db      1245 GGCATTCGCTATGAAACCAAAATGTCTCTTGAGATGACATATCAAAAGAAAGTTGAGAAA 1304
Qy      256 AlaSerIleLysValProGlnLeuLysMetGluIle 267
Db      1305 GTTTCCTAATAAGTACCAATTAATAAATGAGATTA 1340

```

```

RESULT 6
LOCUS   HSA295148      1387 bp      mRNA      linear      PRI 30-OCT-2000
DEFINITION Homo sapiens mRNA for putative sialoglycoprotease type 2.
ACCESSION AJ295148
VERSION  AJ295148.1  GI:11071726
KEYWORDS metallopeptidase; sialoglycoprotease.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS  Chen,J.M., Fortunato,M. and Barrett,A.J.
TITLES   Cloning and sequencing of a second human putative
          sialoglycoprotease homologue
JOURNAL  Unpublished
REFERENCE 2 (bases 1 to 1387)
AUTHORS  Chen,J.M.
TITLES   Direct Submission
JOURNAL  Submitted (27-OCT-2000) Chen J.M., MRC Molecular Enzymology
          Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,
          UNITED KINGDOM
FEATURES
          Location/Qualifiers

```

```

source
1. .1387
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="uterus"
24. .1343
/function="metallopeptidase of family M22"
/codon_start=1
/product="putative sialoglycoprotease type 2"
/protein_id="CAC14666.1"
/db_xref="GI:11071727"
/db_xref="GOA:Q9H4B0"
/db_xref="UniProt/TREMBL:Q9H4B0"
/translacion="MLILTKTAGVFPKSRKRYVERLSFNPFPETLFLKIVIGTET
SCDPTAAVNDPTGVNLSGAIHSQTEVHAKTGIVTPRAQOQIHRNTQIIVQALAS
GVSPDLSIATITIKPGLALISGVLSFSLQVGLKRPPIPHMEALATITLTK
VEPPLVLLISGKCLALVQVSDPFLKGLKLDIAPGMDLRVARRLSLIRKPECT
MSGGALBIHLAKQGRFPHDIKPLPHAKNCDPSFTGLQHTVDKILMKKEGELFI
SKVEQINIPLGLKIAAHCRXYEKQIISADIAATVQHTMACHLVKRTTHAILFCK
ORDILPONNAVIVASGVASNPYIRRALEILTNATQCTLLCPPEPLCTDNGIMIANNG
IERLRGIGLIDIGIRIEPKPIGVDISKEVGASIVPOLKMEI"

ALIGNMENT Scores:
Pred. No.: 5 43e-124      Length: 1387
Score: 1358.50      Matches: 266
Percent Similarity: 91.10%      Conservative: 0
Best Local Similarity: 91.10%      Mismatches: 1
Query Match: 98.09%      Indels: 25
DB: 9      Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x HSA295148 (1-1387)
Qy      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
Db      465 ATGAGAGGCTCATGCACTTAATAAGTTAGTGAACCAATTAAGTAATTCCTTTTAGTT 524
Qy      21 LeuLeuIleSerGlyIleHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      525 CTTTGTATTTCTGGAGGTCACTGTCTGTGGCATTAAGTTCAAGAGATTTTCAGATTTTCG 584
Qy      41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
Db      585 CTTCTTGGAAAAGTCTTTGACATAGCACACAGATGATCATGCTTGAACAAGTGGCAAGAAGA 644
Qy      61 LeuSerLeuIleIysHisProGluCysSerThrMetSerGlyIleLysAlaIleGluHis 80
Db      645 CTTTCTTTAATAAACAATCCAGAGTGTCTCCACCATGAGTGGGGAAGCCATAGACAT 704
Qy      81 LeuAlaIysGlnGlyAsnArgPheHisPheAspIleIysProProLeuHisHisAlaIys 100
Db      705 TTGGCCAAACAAGAAATAGATTTCATTTCATTCATCAACCTCCCTTCATCATCTAA 764
Qy      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
Db      765 AATTGATATTTTCTTTACTGACATTCACACGTTACTGATAAATAATATAAGAAAG 824
Qy      121 GluIysGluGluGlyIle----- 126
Db      825 GAAAAAAGAGAGAGATATTTCTAATTAAGTAAAGTTGAACAGATAATATTCCTGATTTG 884
Qy      127 -----GluIysGlyGlnIleLeuSerSerAla 135
Db      885 TGCCATAAATAAGTCTGCTCATTTCTGCAGGTATGAGAAAGGGCAATCTCTTCACACA 944
Qy      136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155
Db      945 GCAGACATTTGCTGCACAGTACACACACAAATGGCATGTCATCTTGAAAAAACAACAT 1004
Qy      156 ArgAlaIleLeuPheCysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175
Db      1005 CGGGCTATTTCTGTTTGTAAAGCAGAGACTGTGTTACTCTCAAAAATAATGACGTACTGTT 1064

```

QY 176 AlaSerGlyGlyValAlaSerAspPheTyrIleArgArgAlaLeuGluIleLeuThrAsn 195  
DB 1065 GCATCGTGGTGTGCGCAAGTAATCTTATATCCGACAGGCTCTGGAAATTTTAAACAAAC 1124  
QY 196 AlaThrGlnCysThrLeuLeuCysProProProArgLeuGlyThrAspAsnGlyIleMet 215  
DB 1125 GCACACAGTGTGACCTTGTGTGTCTCTCCACACATGACATGACATGAAATGCGATTATG 1184  
QY 216 IleAlaThrPheAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235  
DB 1185 ATTCAGAGGAATGATGATTAAGAAAGACTACGTGTGCTGGCATTTTACATACATAGAA 1244  
QY 236 GlyIleArgTyrGluProLysCysProLeuGlyValAspIleSerLysGlyValGlyGlu 255  
DB 1245 GGCATCGGCTATGACCAAAATGTCCTCTTGAGATGACATATCAAAAGAGTTGAGAA 1304  
QY 256 AlaSerIleLysValProGlnLeuLeuMetGluIle 267  
DB 1305 GCTTCATTAAGATACCAATTAATAATGAGATTA 1340  
RESULT 7  
AX664697 1245 bp DNA linear PAT 22-MAR-2003  
LOCUS AX664697  
DEFINITION Sequence 6 from Patent WO02074960.  
ACCESSION AX664697  
VERSION AX664697.1 GI:29164457  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1  
AUTHORS Leiby, K.R., Kapeller-Libermann, R. and Glucksmann, M.  
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
JOURNAL of human proteins and uses thereof  
Patent: WO 02074960-A 6 26-SEP-2002;  
Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source 1. 1245  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Alignment Scores:  
Pred. No.: 5.14e-123 Length: 1245  
Score: 1348.00 Matches: 260  
Percent Similarity: 98.50% Conservative: 3  
Best Local Similarity: 97.38% Mismatches: 4  
Query Match: 97.33% Indels: 0  
DB: Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x AX664697 (1-1245)  
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
DB 442 ATGAGAGGCTGATGACCTTACTTATAGGTTGACCAATTAAGTAATTTCTTTTATAGTT 501  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 502 CTTTGAATTTCTGAGAGGTCACTGTCTGTGGCATTAAGTTCAAGAGGTTTCAGATTTCG 561  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 562 CTTCTTGAAGAGTCTTGTGACATGACACAGGTGACATGCTTGCACAGGTGCAAGAGA 621  
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
DB 622 CTTTCTTTAATAAATCAATCCAGAGTCTCCACCATGATGTGGGAAAGCCATTAGAACT 681  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
DB 682 TTGGCCCAACAGGAATATGATTTTCACTTTTGACATCAACCTCCCTTGATCATGCTTAA 741

QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
DB 742 AATTGTGATTTTCTTTTACTGACTTCAACACGTTACTGATGATTAATAATATGAAACAG 801  
QY 121 GlnLysGlnGlnGlyIleGlnLysGlnGlnIleLeuSerSerAlaAlaAspIleAla 140  
DB 802 AAACAAGAGGAAGTATGAGAAGGGGCAAAATCTGTCTTCAGCAGCAGCAATTCCTGCC 861  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 862 ACAGTACAGACACACAATGCGATGCTTGTGAAGAAAGAACATCGGCTATTCGTT 921  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValIleValAlaIleSerGlyGlyVal 180  
DB 922 TGTAAAGCAGAGACTGTGTTACTCAAAATATGCAATGCTGCTGTCATCTGGTGTGTC 961  
QY 181 AlaSerAspPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
DB 982 GCAAGTAACTTCTATATCCGACAGCTCGAAATTTTAAACAAACGACACAGTGCACCT 1041  
QY 201 LeuLeuCysProProProArgLeuGlyThrAspAsnGlyIleMetIleAlaThrPheAsnGly 220  
DB 1042 TTGTTGTGTCTCTCCACAGACTATGACATGATTAATGCAATTAATGATGATGATGATG 1101  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240  
DB 1102 ATTGAAAGACTACGTGCTGGCTGGCATTTTACATGACATGAAAGGCACTCCGCTATGAA 1161  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGluAlaSerIleLysVal 260  
DB 1162 CCAAAATGTCTCTTGAGATGACATATCAAAAGAGTTGGAAGACTTCATTAAGTA 1221  
QY 261 ProGlnLeuLysMetGluIle 267  
DB 1222 CCACATTAATAATGAGATTA 1242  
RESULT 8  
AX664695 1820 bp DNA linear PAT 22-MAR-2003  
LOCUS AX664695  
DEFINITION Sequence 4 from Patent WO02074960.  
ACCESSION AX664695  
VERSION AX664695.1 GI:29164455  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1  
AUTHORS Leiby, K.R., Kapeller-Libermann, R. and Glucksmann, M.  
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
JOURNAL of human proteins and uses thereof  
Patent: WO 02074960-A 4 26-SEP-2002;  
Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source 1. 1820  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD80044.1"  
/translation="MLILTKTAGVFFKSKRKYVEFLRSFNHPGTLFLAKIVLGIET  
SCDPRAAAVDSTGNVLEGAHSQTEVHLKTGCIYPPAQOHLREVIQRIVOALASAS  
GVSPSDLSAIAITTKRGALSLGVGLSTGLVGLKKPFIHHEBAALITRLNKK  
VERPVLVILISGHCILALVQSVSPILLGSLDIDARQMDLKVARRSLIISPEKST  
MSGRAIRHIAQGRFHFHDILPPLHAAACDFSTGLDHTVDKNNENKQEBGIEKG  
QILSSAADIAATVQHTMACHLVKRTIRAILFPKQBDLLPQNNAAVLVASSGVASNFYR  
RALRIITAAQCTTLCTPPRLCTDNGIMIAMNGIIRLRAGLGIILHDIIEGIRYEPKCP  
LGVDSKEVGEASIKVPQLMETI"  
CDS  
ORIGIN



## Alignment Scores:

Pred. No.: 8,17e-123 Length: 1820  
 Score: 1348.00 Matches: 260  
 Percent Similarity: 98.50% Conservative: 3  
 Best Local Similarity: 97.38% Mismatches: 4  
 Query Match: 97.33% Indels: 0  
 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x AX664695 (1-1820)

QY 1 MetGluAlaHisAlaLeuThrTlLeaArgLeuThrAsnLysValGlnPheProPheLeuVal 20  
 DB 587 ATGAGGCTCATGACACTACTATTAGTTGACCAATTAAGTAGAATTTCTTTTACTG 646  
 QY 21 LeuLeuLleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
 DB 647 CTTTGTATTTCTGAGGCTGCTGTGTGGCATTAGTTCAAGAGGTTTCAGATTTTCTG 706  
 QY 41 LeuLeuGlyLysSerLeuAspLleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
 DB 707 CTTCTTGAAAGTCTTTTGACATGACACAGGTGACATGCTTGACAGAGTGGCAAGAGA 766  
 QY 61 LeuSerLeuLleLysHisProGlnCysSerThrMetSerGlyGlyLysAlaLleGlnHis 80  
 DB 767 CTTCTTAAATTAATAACATCCAGAGTGTCCACCATGAGTGGTGGAAAGCCATTAGACAT 826  
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspLleLysProProLeuHisAlaLys 100  
 DB 827 TTGGCCAAACAGAGAAATAGATTTTCATTTTGAACATCAACCTCCCTTCATCATCTCTAA 886  
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysLleGlnMetLysLys 120  
 DB 887 AATTGTATTTTCTTTTACTGACCTTCAACACGTTACTGATTAATAAATGAAGAAACAGG 946  
 QY 121 GlnLysGlnGlnGlyLleGlnLysGlnLleLeuSerSerAlaAlaAspLleAlaLle 140  
 DB 947 AAACAAGAGAGATATTGAGAAGGGCAATTCGTCTTGACGACGACGACATTCCTGCC 1006  
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaLleLeuPhe 160  
 DB 1007 ACAGTACAGACACAAATGCAATGTCATCTTGTGAAGAAACACATCGGGCTATTCTGTT 1066  
 QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
 DB 1067 TGTAAAGAGAGAGCTGTGTAACCTCAAAATATGACAGTGTGCACTGTGTGTCTC 1126  
 QY 181 AlaSerAsnPheTyrLleArgArgAlaLeuGlnLleLeuThrAsnAlaThrGlnCysThr 200  
 DB 1127 GCAAGTAACTTCTATATCCGACAGAGCTCTGAAATTTTAAACAAAGCAACACAGTCACT 1186  
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyLleMetLleAlaTyrAsnGly 220  
 DB 1187 TTGTGTGTCTCTCTCCACAGACTATGACATATGACATATGATATGATGATGATGATGAT 1246  
 QY 221 LLeGlnArgLeuArgAlaGlyLysGlyLysLleLeuHisAspLleGlnGlyLleArgTyrGlu 240  
 DB 1247 ATGAAAGACTTACCTGTGCTGTGGCACTTTTACATGACATGACAGAGCATCCGCTATGAA 1306  
 QY 241 ProLysCysProLeuGlyValAspLleSerLysGlnValGlyGlnAlaSerLleLysVal 260  
 DB 1307 CCAAAAGTCTCTTGTGAGTAGCATATCAAAAGAGTTGAGAGGCTTCCATTAAGTA 1366  
 QY 261 ProGlnLeuLysMetGlnLle 267  
 DB 1367 CCACAATTAAATAAGAGATA 1387

RESULT 9  
 BC058172 EC059172 1644 bp . mKNA linear ROD 08-OCT-2003  
 LOCUS M8 musculus cdna clone MGC:67870 IMAGE:5012054, complete cds.  
 DEFINITION BC058172  
 ACCESSION BC058172.1 GI:34849663

## KEYWORDS

## SOURCE

MGC.  
 Mus musculus (house mouse)

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

## AUTHORS

1 (bases 1 to 1844)  
 Strausberg,R.L., Feingold,B.A., Grouse,L.H., Dege,J.G.,  
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,P.,  
 Diatchenko,L., Marusina,K., Farmer,A., Rubin,G.M., Hong,L.,  
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,  
 Carninci,P., Prange,C., Rhee,S.S., Loughran,N.A., Peters,G.J.,  
 Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hu,Y.K., S.W.,  
 Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
 Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,  
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
 Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,  
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
 Butlerfield,Y.S., Krzywinski,M.I., Skalek,U., Smalins,D.B.,  
 Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 MEDLINE  
 PUBMED  
 12477932  
 2 (bases 1 to 1844)  
 Strausberg,R.  
 Direct Submission  
 Submitted (15-SEP-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## JOURNAL

## REMARK

## COMMENT

EMAIL: cgabbs-remail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: nisc.mgc@nih.gov  
 Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,  
 Blakeley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,  
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,  
 Hansen,Q., Ho,S.-L., Karlins,E., Kwong,P., Latic,P., Legaspi,R.,  
 Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,  
 McDowell,J., Pearson,R., Stancirpop,S., Thomas,P.J., Touchman,J.W.,  
 Tsungueon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,  
 Young,A., Zhang,L.-H. and Green,E.D.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LMU at: <http://image.llnl.gov>  
 Series: IRAX Plate: 123 Row: d Column: 22  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 21512463.  
 Location/Qualifiers  
 1. 1844  
 /organism="Mus musculus"  
 /mol\_type="cDNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="MGC:67870 IMAGE:5012054"  
 /ciscn\_type="Mammalian tumor metastasized to lung. Tumor  
 arose spontaneously from a senescent normal mammary  
 (clonal) outgrowth infected with the virus MMTV."  
 /clone\_id="NCI CGAP\_Lu29"  
 /lab\_host="DH10B"

CDS

/note="Vector: pCMV-SPORT6"  
222..1466  
/codon\_start=1  
/product="unknown (protein for MG:67870)"  
/protein\_id="AAH58172.1"  
/db\_xref="GI:34849664"  
/translation="MMLRRTAGAI PKPKSKVYGLRFSVHPRTLSCHKLVIGIET  
SCDPTGAIVDEGTGNVGRALHSOTQVLTQKGIIVPAQOAHRENIORIVETLSAS  
RIPPSDLSAIVTTIKRGLALSIGVLSISLOLVNRFKPPPIPHHMEHALITRIITNK  
VEPPVLVLTISGHCILALVQSDFTLGLGSLDIPQDMDKAKRSLIKHPCST  
MSGKALBHLAKDNRRHFTTNPPQNKQNDPFTGLQHTDKLITKEKEEGLEKG  
QILSSADIIAAVQHAATACILAKRTKRAIILCKQNLISPAVAIVSGVAGSNIYIR  
KALEIVANATQCTLLCPPEPLCTDNGIMIAMNGIERLRAGAGLVLDVEDIRPEPCPL  
GIDISREVAEAAIVRLKML"

misc\_feature  
333..1397  
/note="QRT1; Region: Metal-dependent proteases with  
possible chaperone activity [Posttranslational  
modification, protein turnover, chaperones]"  
/db\_xref="CDD:COG0533"

ORIGIN

Alignment Scores:  
Pred. No.: 2,76e-111 Length: 1844  
Score: 1231.00 Matches: 233  
Percent Similarity: 92.88% Conservative: 15  
Best Local Similarity: 87.27% Mismatches: 19  
Query Match: 88.88% Indels: 0  
DB: Gaps: 10

US-10-649-273-2\_COPY\_148\_414 (1-267) x BC058172.1 (1-1844)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLeuValGluProPheLeuVal 20  
DB 663 ATGAGAGGCTCAGCAGCTAGTAATTAAGGCTCAACAAATAAGAAATTCCTTTCTAGTT 722  
QY 21 LeuLeuIleSerGlyGlyIleCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 723 CTTTGTATTTCTGGCGGTCCTGCTGCTGTTGGCATTAAGTCCAAAGTGTTCCGATTTCTCG 782  
QY 41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspLeuValAlaArgArg 60  
DB 783 CTCCTTGGGAAGTCTTTGACATAGCGCAGCGACATGCTTGAACAAGTGGCAAGAAGA 842  
QY 61 LeuSerLeuIleValHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
DB 843 CTTTCTTTAACTCAACATCAGAAATGTTCTCAATGAGTGGTGAAGCAATATAGAAAT 902  
QY 81 LeuAlaIleGlnGlyAsnArgPheHisPheAspIleLeuProPheLeuHisAlaIleVal 100  
DB 903 TTGGCCAAAGACGAAATAGATTCATTTTACTATCAATCCACTATGACAGATCTAAG 962  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLeuIleIleMetLeuVal 120  
DB 963 AATTGCATTTTCTTTTACGGGACTTCAACATATTATCGAATAGCTATATAACACAG 1022  
QY 121 GluIleGluGluGlyIleGluIleGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
DB 1023 GAAAAAAGAGGAGCATTTGAGAGGGGCAAAATCCGTGATCAGCTGACGACATTCCTCT 1082  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIleValSerGlnHisArgAlaIleLeu 160  
DB 1083 GCGGTACAGCATGCAACAGCGTGCACCTTGCAGAAAAGAACACATCGTGTCTGTGTT 1142  
QY 161 CysIleGlnArgAspLeuLeuProGlnAsnAlaValAlaLeuValAlaSerGlyVal 180  
DB 1143 TGCAAGCAGAAAAATTTGCTATTCCTCAAGCTAAACGAGATTAAGTGTATCTGAGAGTGT 1202  
QY 181 AlaserAsnPhetYrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
DB 1203 GCAAGTAATCTGTGATCCGAAAGCATTTGCGAAATATGTCGCAAAATGACAGACGTGCACG 1262  
QY 201 LeuLeuCysProProPheArgLeuCysThrAspAsnGlyIleMetIleAlaIleTrpAsnGly 220

DB 1263 TTGCTGTGTCCCTCTCAAGACTGTGACTGCAACATGACATCATGATTCAGTAAGTA 1322  
QY 221 IlleGluArgLeuArgAlaGlyLeuGlnIleLeuHisAspIleGlnGlyIleArgTyrGlu 240  
DB 1323 ATTTGAAAGATTACGTCGTCGCTTGGCTTTTACATGATGATGAGAAACATCCGATTAGA 1382  
QY 241 PolyS(CysProLeuGlyValAlaIleSerIleSerGlyValAlaGlyValAlaSerIleVal 260  
DB 1383 CCAAAATGTCTCTTGGATATACATATTCACAGAAAGATTGCAAGACTGCCATTAAGA 1442  
QY 261 ProGlnLeuValMetClnIle 267  
DB 1443 CCGCATTAATAATGACACTT 1463

RESULT 10  
BC038910  
LOCUS  
DEFINITION  
MUS musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA  
clone IMAGE:5053559), partial cds.  
BC038910  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1017)  
Strausberg, R.L., Felingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Schenker, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schenker, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brannstrom, M.J., Ustin, T.B., Toshynski, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettleman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E.,  
Schnerich, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT

2 (bases 1 to 1017)  
Strausberg, R.  
Direct Submission  
Submitted (25-OCT-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcsc.bc.ca](mailto:info@bcsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Boeder, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Gardlan, Ken Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McEavy, Steven  
Nees, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Diane Smilus, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Teai, Natasja van den Bosch, Jill Vardy,

George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLNC at: <http://image.llnl.gov>  
Series: IRAX Plate: 86 Row: E Column: 12.  
Location/Qualifiers

FEATURES

Source  
1. 1017  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5053559"  
/tissue\_type="Liver, normal, 5 month old male mouse."  
/clone\_id="NCI CGAP L19"  
/lab\_host="MDH10B"  
/note="Vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:

Pred. No.: 6,53e-111 Length: 1017  
Score: 1224.00 Matches: 233  
Percent Similarity: 92.51% Conservative: 14  
Best Local Similarity: 87.27% Mismatches: 20  
Query Match: 88.38% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x BC038910 (1-1017)

1 MetGluAhiAalaleuThrlaargleuThraenlyValGluPheProPheleuVal 20  
26 ATGAGAGCTCAGCAGCTATTAAGCTCAACATTAAGATTTCTTTTAAATT 85  
21 LeuLeuileSerGlyYhiScyleuLeuAlaleuValGlnGlyValSerAphleu 40  
86 CTTTGAATTTCTGGCGGTCTGCTGCTGTCATTAAGTCGAAGTGTTCGATTTCTG 145  
41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaarg 60  
146 CTCTTTGGAGTCTTTGGACATAGCACAGCGCATGCTTGAAGAGTGGCAAGACA 205  
61 LeuSerLeuileLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
206 CTTTCTTAATCAACAATCCAGAAATGTTCTAATGAAGTGGGAAAAGCTTATGACAG 265  
81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLys 100  
266 TTGGCCAAAGACGGAATAGATTCATTTTACTTACTCAATCCATCCATGACAGATCTG 325  
101 AsnCysAspPheSerPheThrcGlyLeuGlnHisValThraAspLysIleIleMetLysLys 120  
326 AATTGCGAATTTCTTTCACGCGACTTCACATATTAAGTGAATCAATTAACACACAAG 385  
121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaLys 140  
386 GAAAAAAGAGAGCATTTGAGAGAGGGAATCTCTGATCATCAGCTGCAGACATTCCTG 445  
141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
446 GCGGTACAGCATGCAACAGCGTGCACCTTGCAGAAAAGAACATCGGCTATTCGT 505  
161 CysLysGlnAlaGargPheLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGlyVal 180  
506 TGCAAGAGAAAATTTGCTCTCTCCAGCTAACCGAGATTAAGTGTATCTGAGAGTGT 565  
181 AlaserAspPheYrIleArgAlaLeuGlnIleLeuThraAsnAlaThrcGlnCysThr 200  
566 GCAAGTAACTTGTACATCCGAAGCATTTGGAATTTGCGCAAAATGCAACGCAAGTGCAG 625  
201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleAsnGly 220  
626 TTGTGTGTCACTCCAGAGCTGTGCACTGACATGCAATGCAATGCAATGCAATGCAATG 685  
221 TleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGlyTru 240

|||||.....|||:|||||  
686 ATTGAAAGATACGTCCTGGCTTGAACATGATGTAGAGACATCCGATATGAA 745

241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysVal 260  
746 CCAAAATGTCCTCTTGAAGTACATATCCAGAGAGTTCAGAGAGTTCAGAGAGTTCAGAG 805

261 ProGlnLeuLysMetGluIle 267  
806 CCGCATTTAAATATGCACTT 826

RESULT 11  
AX713716 2208 bp DNA linear PAT 15-APR-2003  
LOCUS Sequence 400 from Patent EP1293569.  
ACCESSION AX713716  
VERSION AX713716.1 GI:29888642  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
AUTHORS  
1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J., I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuho, Y.  
Full-length cDNAs  
Patent: EP 1293569-A 400 19-MAR-2003;  
Helix Research Institute (JP) ; Research Association for Biotechnology (JP)

FEATURES

Source  
1. 2208  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Alignment Scores:  
Pred. No.: 1.57e-108 Length: 2208  
Score: 1204.00 Matches: 239  
Percent Similarity: 89.51% Conservative: 0  
Best Local Similarity: 89.51% Mismatches: 4  
Query Match: 86.93% Indels: 24  
Gaps: 1

US-10-649-273-2\_COPY\_148\_414 (1-267) x AX713716 (1-2208)

1 MetGluAhiAalaleuThrlaargleuThraenlyValGluPheProPheleuVal 20  
785 ATGAGAGCTCAGCAGCTATTAAGTGAACCAATTAAGATTTCTTTTAAATT 844  
21 LeuLeuileSerGlyYhiScyleuLeuAlaleuValGlnGlyValSerAphleu 40  
845 CTTTGAATTTCTGGAGTCTGCTGCTGTCATTAAGTCAAGAGATTTCTG 904  
41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaarg 60  
905 CTCTTTGGAAGTCTTTGACATAGCACAGGTGACATGCTTGAACAGTGGCAAGAGA 964  
61 LeuSerLeuileLysHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
965 CTTCTTAATTAACATCCAGAGTCTCCACATGATGAGTGGGAAACCATTAAGACAT 1024  
81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLys 100  
1025 TTGGCCAAAGAGAAATTAAGATTTCAATTTGACATCAACCTCTTGCATCATGCTTAA 1084  
101 AsnCysAspPheSerPheThrcGlyLeuGlnHisValIleAspLysIleIleMetLysLys 120  
1085 AATTGGAATTTCTTTTATCTGCACTGCAACGTTTACTGATTAATTAATTAATGAAG 1144  
121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaLys 140

```

Db      1145 GAAAAAGGAGGATTTAGAGAGGCGAAATCTGTCTTCAGCAGACATGTCGCC 1204
Qy      141  ThrValGlnHisMetAlaCysHisLeuValIlyArgThrHisArgAlaIleuPhe 160
Db      1205 ACAGTACGACACACATGATGATCTTTGTGAAAAGAACACATCGGCTATTCGT 1264
Qy      161  CysIsglnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerIsglyVal 180
Db      1265 TGTAAAGCAGAGAGACTTGTACTCTAAATATATGACGATCTGTCATCTGTGTGC 1324
Qy      181  AlAsenAspMetIlyIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
Db      1325 GCAAGTAACTTCTGTATCTCGACAGCTCTGGAAATTTTAAACAAGCACAAGTCAGCT 1384
Qy      201  LeuLeuCysProProProArgLeuLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220
Db      1385 TTGTTGTCTCTCTCCACGACTATGACCTGATTAATGACATTATATATGCA----- 1435
Qy      221  IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
Db      1435 ----- 1435
Qy      241  ProLysCysProLeuGlyValAspIleSerIlyGlyValGlyAlaSerIleIysVal 260
Db      1436 ---TGATGTCTCTTGAGATGACATATCAAAAGAGTTGAGAGAGCTTCCTTAAAGTA 1492
Qy      261  ProGlnLeuLysMetGluIle 267
Db      1493 CCACATTAATAAATGAGATA 1513

RESULT 12
AK055441      2208 bp      mRNA      linear      PRI 30-JAN-2004
LOCUS      AK055441
DEFINITION      Homo sapiens cDNA FLJ30879 f18, clone FEBRA2004592, highly similar
                to Homo sapiens mRNA for putative sialoglycoprotease type 2.
ACCESSION      AK055441
VERSION      AK055441.1 GI:16550166
KEYWORDS      oligo capping; f18 (full insert sequence).
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS      Oca,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
                Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
                Sekine,M., Odayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
                Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
                Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
                Shiratori,A., Sudo,H., Hosori,T., Kaku,Y., Kodaira,H., Kondo,H.,
                Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
                Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
                Tanikawa,M., Yamazaki,M., Niomiya,K., Ishibashi,T., Yamashita,H.,
                Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
                Hirooka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
                Yoshida,M., Hottu,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
                Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
                Takeuchi,K., Arita,M., Imose,N., Mueshino,K., Yuki,H., Oshima,A.,
                Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
                Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
                Teraashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizuguchi,H.,
                Goto,Y., Shikizu,F., Wakebe,H., Hishiguchi,H., Watanabe,H.,
                Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
                Kuniagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
                Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
                Oaki,K., Hiraio,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
                Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,K.,
                Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,
                Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togaishi,T.,
                Oyama,M., Hara,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
                Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
                Nagase,T., Nomura,N., Kikuchi,H., Maehuo,Y., Yamashita,R.,
                Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.

```

```

TITLE      Complete sequencing and characterization of 21,243 full-length
JOURNAL      human cDNAs
PUBMED      Nat. Genet. 36 (1), 40-45 (2004)
14702039
REFERENCES
AUTHORS      1 Nishi,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
                Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
                Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
                Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
                Kikuchi,H., Kanda,K., Wagatsuma,M., Murakami,K., Kanehori,K.,
                Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
                Sugano,S., Nagahari,K., Maehuo,Y., Nagai,K. and Isogai,T.
                NEDO human cDNA sequencing project
                Unpublished
                3 (bases 1 to 2208)
REFERENCE
AUTHORS      Isogai,T., Otsuki,T. and Sugiyama,T.
TITLE      Direct Submission
JOURNAL      Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
                Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
                (E-mail:genomics@hri.co.jp, Tel:81-438-52-33975, Fax:81-438-52-3986)
                NEDO human cDNA sequencing project supported by Ministry of
                Economy, Trade and Industry of Japan; cDNA full insert sequencing:
                Research Association for Biotechnology (RAB); cDNA library
                construction: Helix Research Institute (HRI) (supported by Japan
                Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
                HRI, and Biotechnology Center, National Institute of Technology and
                Evaluation; clone selection for full insert sequencing: RAB and
                HRI.

FEATURES
source      Location/Qualifiers
            1..2208
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="FEBRA2004592"
                /tissue_type="brain"
                /clone_id="FEBRA2"
                /dev_stage="fetus"
                /note="cloning vector: pME186FL3"
            344..1438
                /note="unamed protein product"
                /codon_start=1
                /protein_id="BAB70923.1"
                /db_xref="GI:16550167"
                /translation="MLILTKTAGVFPKPSRKRYEFLRSFNFPGTLFLHKIVLGIET
                SDDPRAAVDPETGVNLGEALHSQTEVHLKRGIVPRAOCLHRENIQIYDEALASAS
                GVSPEDSLAIATTPRGLALSLGVGLSRLQVGVQKKEFTPIHMEBALPILRLTNK
                VEPPLVLIISGHLALVQGVSDFLGKSLDLPQMDLKVARRPLIKRPECST
                MSGGAIEHLAQGRFHFIDIKRPLHAKNCDFPGIQLVTDKLIIMKEKEGIEG
                QILSSAADIATVQTMACHIVKRTFRALIFPCQRDLIPONNAVLVAGSVANFCIR
                RALRLTNATQCTLILCPRLCTDNGIMIA"

ORIGIN
Alignment Scores:
Pred. No.:      1,57e-108      Length:      2208
Score:          1204.00      Matches:      239
Percent Similarity: 89.51%
Best Local Similarity: 89.51%      Mismatches:      4
Query Match:    86.93%      Indels:      24
DB:              9      Gaps:      1

US-10-649-273-2_COPY_148_414 (1-267) x AK055441 (1-2208)
Qy      1  MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIlyValGluPheProPheLeuVal 20
Db      785  ATGAGAGGCGTCACTGACTTATGTTAGTGGACCAATTAAGTGAATTCCTTTTATGAT 844
Qy      21  LeuLeuIleSerGlyHisCysLeuLeuAlaLeuValGlnGlyValAspSerPheLeu 40
Db      845  CTTTGTGATTTCTGAGAGTCACTGTCTGTGGCACTTATGTTCAAGAGATTTTCAGTTTCG 904
Qy      41  LeuLeuGlyIlySerLeuAspIleAlaProGlyAspMetLeuAspIlyValAlaArgArg 60
Db      905  CTTTGTGAAATCTTTTGGACATAGCACAGGTGACATCTTGCACAAAGTGGCAAAAG 964

```

```

QY      61      LeuSerLeuIleuYSHiSPProGluCySerTherMetSerGlyValAlaIleGluHis 80
      |||
      965      CTTCCTTTAATAAACAATCCAGAGTCTCCACCAATGATGGGAGAAACCTAAGACAT 1024
QY      81      LeuAlaIleGluGlyAaMaArgPheHisPheAspIleLeuYsProProLeuHisIleAlaYs 100
      |||
      1025      TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATCTAAA 1084
QY      101      AsnCyAspPheSerPheTherGlyLeuGlnHisValThrAspYsIleIleMetLeuYs 120
      |||
      1085      AATTGTATTTTCTTTTACAGCTTCAACACCTTACTGTAAATATATATGAAAAAG 1144
QY      121      GluYsGluGluGlyIleGluYsGlyValIleLeuSerSerAlaAlaAspIleAlaIle 140
      |||
      1145      GAAAAAAGAGAGATATTGAGAAAGGGCAATCTCTGTTCCAGCAGCAGCATCTCTGCC 1204
QY      141      ThrValGlnHisTherMetAlaCyHisIleuValIlyAsArgThrHisArgAlaIleLeuPhe 160
      |||
      1205      ACAGTACAGACACAAATGGCATGTCTTGTGAAAAAGAACATCGGGCTATTTCTGTTT 1264
QY      161      CyAluGluGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGlyVal 180
      |||
      1265      TGTAAAGCAGAGAGACTTGTTAACCTCAAAATATGACGTAAGTGTGCTGTGTGTCTC 1324
QY      181      AlaSerAsnPheTyrlIeArgArgAlaIleuGlnIleLeuThrAspAlaIleTherGlnCyThr 200
      |||
      1325      GCAAGTAACCTTCTGTATTCGAGAGCTTCGAAATTTTAACAAACGCAACACAGTGCACT 1384
QY      201      LeuLeuCyAspProProArgLeuCyThrAspAsnGlyIleMetIleAlaIleArgAsnGly 220
      |||
      1385      TTGTGTGTCTCTCTCCAGCATATGCACTGATTAATGCACTTAATGATTGCA----- 1435
QY      221      IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrlu 240
      |||
      1435      ----- 1435
QY      241      ProIleCyAspProLeuGlyValAlaAspIleSerIleGlyValAlaSerIleLeuVal 260
      |||
      1436      --TGAATGCTCTCTGGAGTAGACATATCAAAAGAAAGTTGAGAAAGCTTCCATAAAGTA 1492
QY      261      ProGlnLeuIleuYsMetGluIle 267
      |||
      1493      CCACATTAATAATGAGATTA 1513
Db
QY      1435      ----- 1435
QY      241      ProIleCyAspProLeuGlyValAlaAspIleSerIleGlyValAlaSerIleLeuVal 260
      |||
      1436      --TGAATGCTCTCTGGAGTAGACATATCAAAAGAAAGTTGAGAAAGCTTCCATAAAGTA 1492
QY      261      ProGlnLeuIleuYsMetGluIle 267
      |||
      1493      CCACATTAATAATGAGATTA 1513
Db
RESULT 13
LOCUS      BC078974      1546 bp      mRNA      linear      ROD 03-AUG-2004
DEFINITION      Rattus norvegicus cDNA clone IMAGE:7111906, partial cds.
ACCESSION      BC078974
VERSION      BC078974.1 GI:50926879
KEYWORDS
SOURCE
ORGANISM      Rattus norvegicus (Norway rat)
              Mammalia; Euteleostomi;
              Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
              1 (bases 1 to 1546)
REFERENCE
AUTHORS      Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
              Klausner, R.D., Collins, P.S., Wagner, K.H., Shenmen, C.M., Schuler, G.D.,
              Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
              Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
              Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
              Stachenko, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,
              Schetz, T.E., Brownstein, M.J., Udell, T.B., Tosliyski, S.,
              Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
              Abramson, R.D., Mulvaney, S.J., Bosek, S.A., McEwan, P.J.,
              McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
              Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huijy, S.W.,
              Viatelson, D.R., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
              Fahey, J., Hellon, B., Kettman, M., Madan, A., Rodriguez, S.,
              Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shechenko, Y.,
              Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, B.D.,

```

```

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
JOURNAL
TITLE
REMARK
COMMENT
Email: cgabbs-remail.nih.gov
Tissue Procurement: Howard Jacobs
CDNA Library Preparation: Express Genomics
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@pax11.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAX Plate: 182 Row: F Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
1. 1546
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7111906"
/tissue_type="Testis, rat (Brown Norway)"
/clone_1lb="N1H MGC_237"
/lab_host="DH10B"
/note="vector: pExpress1"
144..>1546
/codon_start=1
/product="Unknown (protein for IMAGE:7111906)"
/protein_id="AAH78974.1"
/db_xref="GI:50926880"
/translation="MTLSTKTAGAIIPRRSVRGIRRPVQQLHRENTORIIEBALAS
SCDTRAAVDEBGNVLGBALSHQTEVHAKTGIVPVAAQLRHNTORIIEBALAS
GVSPDLIAITTKPGALSLGVLSFVQVLPKPKPIPIHNEALALTRLTHK
VGPPELVLLISGHCILALVQSDFLIGKSLDIAFGMDLKVARSLLIGHPECT
MSGGAIVHIAKEGRFPETINPQMANCDPFSFGQHTVDKLTGKEGIEKG
OILSSGADIIAAVOATACHLAKRTHRALIPCOQNLVSPAAVIVVSGVSNLYIR
RALRIVANAOTCLCPRLCTDNGIMIANGRIRLRPGSLGIDHVEDIRPERPAGE
IVSWLRVLTALTEDVSSTHTVAHPLNSGRANITQTSWCSCTQIVRTVRLTL
NINLEKSKKKKKKKKK"
ORIGIN
Alignment Scores:
Pred. No.: 4.7e-100 Length: 1546
Score: 1116.00 Matches: 214
Percent Similarity: 92.56% Conservative: 10
Best Local Similarity: 88.43% Mismatches: 18
Query Match: 80.58% Indels: 0
DB: 10 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x BC078974 (1-1546)
QY      1      MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGluPheProPheLeuVal 20
      |||
      585      ATGGAGGCTCAGCCCTCACTATTGAGCTGACCCACAAAGTCGATTTCTTTTATGATT 644

```

QY 21 LeuLeu1eSerGlyGlyHisCysLeuLeu1aLeuValGlnGlyValSerAspPheLeu 40  
Db 645 CTTTGTGATTTCTGAGGCGCACTGCTGTGCGCTTACGTTCAGAGTTCAGATTTTCTG 704  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 705 CTCCTCGGGAAGTCCCTGAGCATAGCCGCGACGACATGCTTGACAAAGTGCGACAGAAGA 764  
QY 61 LeuSerLeuIleLysHisProGlnCysSerThrMetSerGlyGlyValAlaIleGlnHis 80  
Db 765 CTTTCTTTAACTCAAACTCCAGAAATGTTCTACAAATGATGTGTGGAAAGCTATTAACAAT 824  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
Db 825 TTGGCCAAAGAGAAATAGATTCCACTTTACTTACTATCAATCCACCATGACAAATCTTAAG 884  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 885 AACTGTATTTTCTTTTACGGAAGCTTCAACATGTCACCGATTAAGCTATTAACACACAAG 944  
QY 121 GlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
Db 945 GAAAAAAGAGAGGCAATGAGAGAGGCGCAATCTGTATCATCAGCGCAGACATTCCTCT 1004  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
Db 1005 GCGGTATACGACGACAAAGCGCTGCCACTTTCGAAAAGAACATCGTCTATTTCTGTTT 1064  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
Db 1065 TGCCAGCAGAAATTTGCTATCTCCAGCTAACGCAAGATTAAGTTGTGTGCGAGGTGT 1124  
QY 181 AlaSerAspPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
Db 1125 GCAAGTAACTGTATCATCCGAAAGCAATGGAATTTGAGCAATGCAACCAATGCACT 1184  
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220  
Db 1185 TTGTTGTGTCCCTCCGAGCTGTGACTGACATGACAAATGATCATGATGATGCAATGCA 1244  
QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrgln 240  
Db 1245 ATGAAAGATTAAGCTGTGCTGCTGGGCAATTTAATGATGATGATGAAGACATCGATACGA 1304  
QY 241 ProLys 242  
Db 1305 CCAAG 1310

RESULT 14  
BC051211

LOCUS BC051211 1109 bp mRNA linear ROD 15-APR-2003  
DEFINITION Mus musculus, clone IMAGE:1327545, mRNA.  
ACCESSION BC051211  
VERSION BC051211.1 GI:29881634  
KEYWORDS Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Strausberg, R.  
1 (bases 1 to 1109)  
Direct Submission  
Submitted (14-APR-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Marcello Bento Soares, Ph.D.  
cDNA Library Preparation: M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>  
contact: [amadan@systemsbio.org](mailto:amadan@systemsbio.org)  
Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 113 Row: b Column: 1  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.  
FEATURES  
source location/Qualifiers  
1..1109  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1327545"  
/tissue\_type="Thymus gland, mouse"  
/clone\_id="Soares\_thymus\_2bwmr"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac"  
ORIGIN  
Alignment Scores:  
Pred. No.: 6,236-84 Length: 1109  
Score: 950.50 Matches: 190  
Percent Similarity: 79.10% Conservative: 22  
Best Local Similarity: 70.90% Mismatches: 35  
Query Match: 68.63% Indels: 21  
DB: 10 Gaps: 4  
US-10-649-273-2\_COPY\_148\_414 (1-267) x BC051211 (1-1109)  
QY 18 PheLeuValLeuLeuIleSer-----GlyGlyHisCysLeu----- 29  
Db 82 TTTTGTGATTTACGATTCGCGATGCGGTGCTGCAATTTTAACCATCCCTATATCTCT 141  
QY 30 -----LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAsp 47  
Db 142 TCTGTGTGAACTCACTGTCATATGAGGCTTATCTTTGCTTAAT-----AAT 192  
QY 48 IleAlaProGlyAspMetLeuAspLys-----ValAlaArg 59  
Db 193 TTGCAGCCAAATATATTTATCAAAAATAATGATTCCTTTTAAATAGTGCGCAAG 252  
QY 60 ArgLeuSerLeuIleLysHisProGlnCysSerThrMetSerGlyGlyValAlaIleGln 79  
Db 253 AGACTTTCTTTATCAAAATCATCCAGATGTTCTAACAATGATGATGCGAAAGCTATAGAA 312  
QY 80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAla 99  
Db 313 CAGTTGGCCAAAGAGAGCAATGATGATTCATTTACTATCAATCCACTATGCAAGATGCT 372  
QY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119  
Db 373 AAGAAATGCGATTTCTTTACCGGACTTCAACAAATATACGATAGATGCTATTAACACAC 432  
QY 120 LysGlnLysGlnGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 139  
Db 433 AAGGAAAAAGAGAGAGGATGAGAGAGGCGCAATCTGTATCATGCTCAGACATTTGCT 492  
QY 140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159  
Db 493 GCTGGGTATCAACATGCAACAGCGGCGCACTTTCGAAAGAAACATGCGCTATTTCTG 552  
QY 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGly 179  
Db 553 TTTTGCAGCAGAAATTTGCTCTCTCCAGCTTAACGATGATGATGATGATGAGAGG 612  
QY 180 ValAlaSerAspPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199  
Db 613 GTTGCAGATTAATTTATCATCCGAAAGCATTTGGAATTTGTGCAATTCACACGCACTGC 672

QY 200 ThrleuEuCySPProProArgleuCyThrAspAsnGlyIleMetIleAlaTrpAsn 219  
Db 673 ACCTGTTGTGTCTCCACTCCAGACTGTGCATGCATGCATCATGATTCGATGAAAT 732  
QY 220 GlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyr 239  
Db 733 GGAATTTGAAAGATATCGTGTCTGGCTTGGCGCTTTTACATGATGTAGAAAGACATCCGATAT 792  
QY 240 GluProCysPheProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLys 259  
Db 793 GAACCAAAATATCTCTTGTGAGTAGACATATCCAGAGATTGCAAGAGCTCCATTAATA 852  
QY 260 ValProGlnLeuLysMetGluIle 267  
Db 853 GTACCCGATTAAAAATGCACTT 876  
RESULT 15  
BX934991 1522 bp mRNA linear VRT 02-FEB-2004  
LOCUS BX934991  
DEFINITION BX934991 Gallus gallus finished cDNA, clone CHEST189114.  
ACCESSION BX934991  
VERSION BX934991.1 GI:41635519  
KEYWORDS  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 1522)  
Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Gardner, C., Chalk, S.E.,  
Croning, M.D.R., Davies, R.M., Francis, M.D., Graham, D.V.,  
Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,  
Mblet, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,  
Tickle, C. and Wilson, S.A.  
Direct Submission  
Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: chickens@hms.unist.ac.uk  
BBRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA  
sequencing project.  
This sequence is from the  
BBRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
from a library constructed by Elizabeth Boech. cDNA was prepared  
from RNA extracted from muscle, normalised, and poly A-tailed.  
ECORI-NotI cut cDNA was then ligated into the vector. Vector:  
pBlueScript II KS(+); Site\_1: EcoRI; Site\_2: NotI Host: Escherichia  
coli DH10B.  
FEATURES  
source location/Qualifiers  
1..1522  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Layer and broiler"  
/db\_xref="taxon:9031"  
/clone="CHEST189114"  
/clone\_lib="CSBQREN11"  
/dev\_stage="adult"  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,24e-82 Length: 1522  
Score: 939.00 Matches: 176  
Percent Similarity: 81.06% Conservative: 38  
Best Local Similarity: 66.67% Mismatches: 50  
Query Match: 67.80% Indels: 0  
DB: 5 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x BX934991 (1-1522)  
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 587 ATGAGGCTACGCACTTACATCAGATGACAGAGCAAGTAGAATTTCCCTTCTAGTT 646  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40

Db 647 CTTTACTCTCCGAGAGTCACTGCATCTTGGCAGTAGACAGGAGATTTCAGATTCCCT 706  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaGlyArg 60  
Db 707 CTGCTTGACAGCTCCATTAATATAGCACAGGAGATGCTGTGATAGGTAAGCAAGAG 766  
QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80  
Db 767 CTCTCTTTGAGAGCACTCCGAGAGTCCACGGAGATGCTGGGGAGAGGCAATAGAGAC 826  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLys 100  
Db 827 CTGGCTCAACCGAGAGACTGGCAACAGTACCTTGACCTTCCATGCAACAGTATCGT 886  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 887 AACTGTGATTTTCTTCTCCGAGCTTCAGAGCTTCAGAACCAACCAATTCCTTCAGAAA 946  
QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaIleAspIleAla 140  
Db 947 GAAAGAGAGAGAGTATTCAGAGAGGAGAAATCTGTCTGCTTAAGACATCGCTGCT 1006  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
Db 1007 GCTGCACAGACAGTAGCTGCTGCTCATATTATTCACGAGACACAGGACCATGCTTTC 1066  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180  
Db 1067 TGCATGAAGAAACAGCATATTATTACCAAAACCTGCAACTGTGTGTATTCAGAGAGTT 1126  
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
Db 1127 GCAAGTAACAGTATATATCGAAAGAGACCTGCAACTCTGCAAAATCCGATTTTGGT 1186  
QY 201 LeuLeuCysProProProArgleuCyThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
Db 1187 TTTCTGTCTCTCTCCCTCCAGGCTGTGCACCGATATGATGTATGATGATGGAATGGC 1246  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240  
Db 1247 ATTTGAAAGTTTCGTCAGAGATGTGTATTTTATACAGTACGATGCGCTTCGAA 1306  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
Db 1307 CCAAAAGCTCCCTTGTGATGATATTTCCAAAGATTGAAGAGATTCATCAAAAGTG 1366  
QY 261 ProGlnLeuLys 264  
Db 1367 CCAGACTTAAGG 1378

Search completed: February 16, 2005, 18:02:51  
Job time : 4263.78 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:09:32 ; Search time 508.461 Seconds  
(without alignments)  
3108.540 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHLTIRLTNKEVPEPLV.....DISKEVGASIKVPLQKMEI 267

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 segs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seg length: 0  
Maximum DB seg length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+g2n.model -DEV=x1h  
-Q=/cgn2\_1/USPTO.spool/US10649273/runat\_14022005\_114702\_16389/app\_query.fasta\_1.1429  
-DB=N\_Geneseq\_16Dec04 -QMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOFC=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomum62 -TRANS=human40.cdi  
-LIST=45 -POCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR=US10649273 @CGN\_1\_1\_1057 @runat\_14022005\_114702\_16389 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGOUDERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_16Dec04:  
1: geneseq1980s:  
2: geneseq1990s:  
3: geneseq2000s:  
4: geneseq2001as:  
5: geneseq2001bs:  
6: geneseq2002as:  
7: geneseq2002bs:  
8: geneseq2003as:  
9: geneseq2003bs:  
10: geneseq2003cs:  
11: geneseq2003ds:  
12: geneseq2004as:  
13: geneseq2004bs:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	100.0	1416	8	ABX70950 Novel hum
2	1385	100.0	1526	6	ABX76639 DNA encod
3	1385	100.0	2058	6	ABA93268 Human O-s
4	1385	100.0	2197	6	ABX76635 DNA encod
5	1385	100.0	2572	8	ABT23207 Human pro

6	1348	97.3	1820	6	AAD46856	AAD46856 Human g1y
7	1348	97.3	1820	8	ACA60887	ACA60887 Human cdn
8	1348	97.3	1821	10	ABX57020	ABX57020 CDNA enco
9	1204	86.9	2208	10	ADA52832	ADA52832 Human cod
10	1204	86.9	2890	12	ADQ24627	ADQ24627 Human bot
11	995.5	71.9	3358	10	ADB31345	ADB31345 Human dia
12	983	71.0	1572	6	ABQ75508	ABQ75508 Murine s1
13	599	43.2	2734	5	ABQ84622	ABQ84622 DNA encod
14	468	33.8	371	12	ADL86725	ADL86725 DNA up-re
15	468	33.8	371	12	ADL86726	ADL86726 DNA up-re
16	415.5	30.0	1601	4	ABL24633	ABL24633 Drosoph1l
17	415.5	30.0	3656	4	ABL24632	ABL24632 Drosoph1l
18	401.5	29.0	1557	3	ACA38454	ACA38454 Arabidops
19	342	24.7	4360	6	AAD48239	AAD48239 Ehrlichia
20	338.5	24.4	1385	4	AAH15110	AAH15110 Human cdn
21	321.5	23.2	1146	8	ACA26804	ACA26804 Prokaryot
22	314.5	22.7	936	8	ACA20445	ACA20445 Prokaryot
23	308.5	22.3	1044	8	ACA39102	ACA39102 Prokaryot
24	308.5	22.3	94750	4	AAF28551	AAF28551 Genomic f
25	307	22.2	1053	12	ADL03120	ADL03120 DNA encod
26	304.5	22.0	1032	8	ACA43173	ACA43173 Prokaryot
27	301.5	21.8	1000	4	AAE91424	AAE91424 Moraxella
28	301.5	21.8	1000	6	ABK37804	ABK37804 DNA seque
29	300	21.7	1092	6	ABQ90383	ABQ90383 M. capsul
30	291.5	21.0	1026	4	AA554064	AA554064 Pseudomon
31	291.5	21.0	1026	8	ACA42146	ACA42146 Prokaryot
32	291.5	21.0	1026	10	ADG73341	ADG73341 P aerugin
33	291.5	21.0	1059	11	ABD02280	ABD02280 Pseudomon
34	291.5	21.0	1206	11	ABD02197	ABD02197 P aerugin
35	289.5	20.9	1026	10	ADG73343	ADG73343 P aerugin
36	289.5	20.9	1029	4	AA553309	AA553309 Haemophil
37	289.5	20.9	1029	8	ACA34150	ACA34150 Haemophil
38	289.5	20.9	110000	2	AAE42063_05	AAE42063_05 Continuation (6 of
39	285	20.6	372	5	ABV61069	ABV61069 Human pro
40	282.5	20.4	9967	13	ADT05493	ADT05493 Haemophil
41	282.5	20.4	85814	13	ADT05494	ADT05494 Haemophil
42	280.5	20.3	1020	8	ACA44384	ACA44384 Prokaryot
43	280.5	20.3	1074	10	ADP02056	ADP02056 Bacterial
44	279.5	20.2	1014	4	AA556045	AA556045 Salmonell
45	279.5	20.2	1014	8	ACA51431	ACA51431 Prokaryot

ALIGNMENTS

RESULT 1	ABX70950	ABX70950 standard; cDNA; 1416 BP.	
XX	AC	ABX70950;	
XX	DT	05-MAR-2003 (first entry)	
XX	XX		
DB	DR	Novel human cDNA sequence #175.	
XX	KW	Human; gene; sg; nervous system disorder; peripheral neuropathy;	AAD46856 Human g1y
KW	KW	Huntington's disease; amyotrophic lateral sclerosis; haemophilia;	ACA60887 Human cdn
KW	KW	neurodegenerative disease; Parkinson's disease; Alzheimer's disease;	ABX57020 CDNA enco
KW	KW	autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;	ADA52832 Human cod
KW	KW	insulin-dependent diabetes mellitus; anaemia; thrombocytopenia; wound;	ADQ24627 Human bot
KW	KW	ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;	ADB31345 Human dia
KW	KW	fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;	ABQ75508 Murine s1
KW	KW	coagulation disorder; cancer; tumour; inflammatory disease; septic shock;	ABQ84622 DNA encod
KW	KW	Crohn's disease; anaphylaxis; proliferation; chemotactic;	ADL86725 DNA up-re
KW	KW	differentiation; stem cell growth factor; haematopoiesis; chemokinetic;	ADL86726 DNA up-re
KW	KW	haemostatic; antiinflammatory; expressed sequence tag; EST.	ABL24633 Drosoph1l
OS	XX	Homo sapiens.	ABL24633 Drosoph1l
XX	XX		ABL24632 Drosoph1l
XX	XX	WO200281731-A2.	ACA38454 Arabidops
PD	XX	17-OCT-2002.	ADG73343 P aerugin
XX	XX		ADG73341 P aerugin
PF	XX	29-JAN-2002; 2002WO-US001222.	ADT05493 Haemophil



XX 30-JAN-2001; 2001US-00774528.  
 PR (HYSB-) HYSEQ INC.  
 PA (GOOD/) GOODRICH R W.  
 XX  
 XX Tang TX, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX WPI; 2003-058563/05.  
 DR  
 XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid  
 PT or lymphoid cell disorders, bone disorders, mechanical and traumatic  
 PT disorders, coagulation disorders, and inflammatory diseases.  
 XX  
 PS Claim 1; Page; 612pp; English.  
 XX  
 CC This invention relates to the cDNA sequences encoding an isolated novel  
 CC human polypeptide. The protein encoded by the nucleic acid of the  
 CC invention is useful for treating central and peripheral nervous system  
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic  
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,  
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus  
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)  
 CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)  
 CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,  
 CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head  
 CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;  
 CC bacterial, viral or fungal infections; allergic conditions such as  
 CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);  
 CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's  
 CC disease, anaplasia). The protein may be used to inhibit the growth,  
 CC infection or function of infectious agents such as bacteria, fungi,  
 CC viruses, or to effect bodily characteristics, biorthylms or circadian  
 CC cycles of rhythms. The protein may also have  
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis  
 CC regulation, immune stimulation or suppressing, chemotactic/chemokinetic,  
 CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
 CC activities. The cDNA sequences of the invention are useful for expressing  
 CC recombinant protein for analysis. The present sequence represents a novel  
 CC human cDNA sequence of the invention, this sequence is an expressed  
 CC sequence tag (EST) and was identified using subtractive hybridisation  
 XX  
 XX Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 4,68e-147 Length: 1416  
 Score: 1385.00 Matches: 267  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 8 Gaps: 0  
 DB: 8 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x ABX70950 (1-1416)

QY 1 MetGUAAlAHAlaLeuThrIleArgLeuThrAsnIleValGluPheProhelaVal 20  
 DB 502 ATGAGGGCTCAGACCTTACTATTAGGTTGACCAATTAAGTAATTCCTTTTATGTT 561  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
 DB 562 CTTTGTGATTTCTGGAGGTCACCTGCTGTGGCATTTAGTTCAAGAGCTTTCAGATTTTCTG 621  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
 DB 622 CTTCTTGGAAGAGCTTTTGGACATACACAGGTGACATCTTGACAGGTGACAGAAAGA 681  
 QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyLysAlaIleGluHis 80  
 DB 682 CTTTCTTAAATTAACATCCAGAGGCTCCACCAAGAGGGTGGGAAACCCATAGAACAT 741  
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLys 100

DB 742 TTGGCCAAACAGAAATATGATTTTCATTTTGACATCAACCTCCCTTGATCATGCTAA 801  
 QY 101 AaenCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
 DB 802 AATGTGATTTTCTTTTACTGACTCAACAGCTTACTGATTAATAATTAATGAAG 861  
 QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
 DB 862 GAAAAAGGAGAGGATTTGAGAAAGGGCAATCTCTCTTCAGCAGACATGCTGCC 921  
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 DB 922 ACAGTACACACACATATGATGATGATCTTGGAAGAAACACATCGGGCTATTCTGTT 981  
 QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180  
 DB 982 TGTAAAGCAGAGACCTGTTTACCTCAAAATTAATGACATGCTGTCATCTGTCGTC 1041  
 QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
 DB 1042 GCAAGTAACTTCTATATCCGACAGCTCTGGAATTTTAAACAGCAACAGTGCAT 1101  
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIlePhe 220  
 DB 1102 TTGTTGTGCTCTCCCTCCAGACTGATGCACTGATTAATGACATTAATGATGCAATGAT 1161  
 QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240  
 DB 1162 AATTGAAGCTAGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1221  
 QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlnValAspIleLysVal 260  
 DB 1222 CCAAAATGCTCTCTGAGATGACATATCAAAAGATTGGAGAGCTTCCATTAAGTA 1281  
 QY 261 ProGlnLeuLysMetGluIle 267  
 DB 1282 CCACATTAATAATGAGAGATA 1302  
 DB  
 RESULT 2  
 ABS76639  
 ID ABS76639 standard; DNA; 1526 BP.  
 XX  
 AC ABS76639;  
 XX  
 DT 11-DEC-2002 (first entry)  
 DE  
 XX DNA encoding novel human metalloprotease MP1 fragment #1.  
 DE  
 XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
 KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
 KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 KW neurological disorder; gene; ds.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200272751-A2.  
 PN  
 XX 19-SEP-2002.  
 PD  
 XX 05-FEB-2002; 2002WO-US003353.  
 PF  
 XX 05-FEB-2001; 2001US-0266518P.  
 PR 10-APR-2001; 2001US-0282814P.  
 PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 XX  
 PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;

XX WPI; 2002-723329/78.  
 DR P-PSDB; ABG96487.  
 XX  
 PT New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
 PT treating, or ameliorating diseases associated with aberrant  
 PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
 PT neurological disorders.  
 XX  
 PS Disclosure; Page 462-463; 473pp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC metalloproteinase (MP-1). (I) is useful for preventing, treating, or  
 CC ameliorating a medical condition, particularly an immune disorder, an  
 CC aberrant glutamate transport or motor neuron disorder, such as  
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
 CC condition. The compositions and methods are also useful for diagnosing,  
 CC prognosticating, treating, ameliorating and/or treating disorders  
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease). This sequence represents a  
 CC metalloproteinase MP1 polynucleotide  
 CC  
 SQ Sequence 1526 BP; 484 A; 297 C; 289 G; 456 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 5.2e-147 Length: 1526  
 Score: 1385.00 Matches: 267  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-649-273-2\_COPY 148\_414 (1-267) x ABS76639 (1-1526)

QY 1 MetGuaIaAh;SAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
 DB 1 ATGAGAGCTCAGCTTACTTATAGGTTGACCAATAAGTGAATTTCTTTTAGTT 60  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
 DB 61 CTTTGTGATTTCTGAGGTCACGTGCTGTTGGCATTAGTTCAAGAGTTTCAATTTTCG 120  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
 DB 121 CTTCTTGAAAGTCTTGTGACATGACACCAAGGTGACATCTTGACAAAGTGGCAAGAA 180  
 QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysValAlaIleGlnHis 80  
 DB 181 CTTTCTTAAATAAACATCCAGAGTCTCCACCAAGAGTGGGAAAGCCATAGAACT 240  
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisIleAlaLys 100  
 DB 241 TTGGCCAAACAAAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATCATCTAA 300  
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
 DB 301 AATTGTGATTTTCTTACTGAGCTTCAACAGTTACTGATGAATAATATATGAAG 360  
 QY 121 GluLysGluGluGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIle 140  
 DB 361 GAAAAAGGAGAGATTAATGAGAAAGGCAATCTGCTTCAGCAGCAACATCTGCTCC 420  
 QY 141 ThrValGlnHisIleMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 DB 421 ACAGTACGACACCAATGCAATGATTCATCTTGAAAGAAACATCGGCTATCTGTTT 480

QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyGlyVal 180  
 DB 481 TGTAAACAGAGAGACTTGTACTCTCAAAATAATAGCAGTACTGTTGCATCTGTTGATGTC 540  
 QY 181 AlaSerAspPheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
 DB 541 GCAAGTAACCTTCATATCCGACAGCTCTGGAATTTTAAACCAACACACAGCTCAGCT 600  
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTPAAsnGly 220  
 DB 601 TTTGTGTGCTCTCTCCAGACTATGCACTGATTAAGCAATTAATGATGATGAATGGT 660  
 QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrglu 240  
 DB 661 ATTGAAGACTAGCTGCTGCTGGCATTTTACAGACATTAAGGCAATCCGCTATGA 720  
 QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
 DB 721 CCAAAATGCTCTTGAGATGACATATCAAAAAGAGTTGAGAGACTTCCATTAAGA 780  
 QY 261 ProGlnLeuLysMetGluIle 267  
 DB 781 CCACAAATTAATAATGAGATTA 801  
 RESULT 3  
 ID ABA93268 standard; cDNA, 2058 BP.  
 AC ABA93268;  
 XX  
 DT 19-APR-2002 (first entry)  
 DE Human O-sialoglycoproteinase-like protein encoding cDNA SEQ ID NO:1.  
 XX  
 KW Human; O-sialoglycoproteinase-like protein; OSGPLP; enzyme; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 110..1354  
 FT /tag= a  
 FT /product= "O-sialoglycoproteinase-like protein"  
 XX  
 PN CNI1318550-A.  
 XX  
 PD 24-OCT-2001.  
 XX  
 PF 19-APR-2000; 2000CN-00106834.  
 XX  
 PR 19-APR-2000; 2000CN-00106834.  
 XX  
 PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.  
 XX  
 PI Mao Y, Xie Y;  
 XX  
 DR WPI; 2002-115090/16.  
 DR P-PSDB; ABB05481.  
 XX  
 PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful  
 PT for diagnosing, preventing and treating related diseases.  
 XX  
 PS Claim 5; Page 29-30 (Disclosure); 38pp; Chinese.  
 XX  
 CC The present sequence encodes human O-sialoglycoproteinase-like protein  
 CC (OSGPLP). The present invention also describes: (1) the preparation of  
 CC the OSGPLP protein; (2) applying the OSGPLP protein in diagnosis; (3) the  
 CC prevention and/or treatment of related diseases; (4) utilizing the OSGPLP  
 CC protein in screening its agonist, excitomotor and inhibitor and preparing  
 CC an antibody against the OSGPLP protein; and (5) the use of the OSGPLP  
 CC polynucleotide sequences, proteins, agonists, excitomoters, inhibitors  
 CC and antibodies in treating diseases related to the abnormal OSGPLP gene  
 CC and in preparing the medicine composite for the treatment  
 XX

SQ Sequence 2058 BP; 637 A; 400 C; 410 G; 611 T; 0 U; 0 Other;

**Alignment Scores:**

Pred. No.:	7.97e-147	Length:	2056
Score:	1385.00	Matches:	267
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) X ABA93268 (1-2058)

Qy	1	MetGluAlaHisAlaLeuThrLleArgLeuThrAsnLysValGluPheProPheLeuVal	20
Db	551	ATGAGGCTCATGACCTTACTATTAGTTGAGCCAAATTAAGATGATTTCTTTTGTAGTT	610
Qy	21	LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu	40
Db	611	CTTTTGATTTCTGGAGGCTCATCTGTCGTGTGGCATATTGTTCAAGAGGTTTCAGATTTTCTG	670
Qy	41	LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg	60
Db	671	CTTTGTGAAAGTCTTTGGACATGACACAGGTGACACGCTTGCACAGGTGGCAAGAGA	730
Qy	61	LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHis	80
Db	731	CTTTCTTTTAATTAATAATCCAGAGTCTCTCCACCATGATGTTGGAAAGCCATGACAT	790
Qy	81	LeuAlaLysGlnGlyAsnArgPheHisIleAspIleLysProProLeuHisHisAlaLys	100
Db	791	TTGGCCAAACAGAGAAATAGATTCTCATTTTGACATCAAACTCCCTTGATCATGCTTAA	850
Qy	101	AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys	120
Db	851	AATGTGATTTTCTTTTACTGAGCTTCACACGTTACTGATTAATAATATATGAAAAAG	910
Qy	121	GluLysGluGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla	140
Db	911	GAATAAAGAGAGGATATTGAAAGGGCAAAATCTCTTTCAGCAGCAGCATTTGCTGCC	970
Qy	141	ThrValGlnHisIsthrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe	160
Db	971	ACAGTACAGCACACAAATGGCATGTGCATCTTGTGAAAAAACAACATCGGGCTATTCTGTTT	1030
Qy	161	CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGlyVal	180
Db	1031	TGTAAAGCAGAGAGCTTGTAACCTCAAAATATGACGATCGTGTGCATCGTGGTGTCTC	1090
Qy	181	AlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr	200
Db	1091	GCAAGTAACTTCTTATATCCGACAGGCTGTGAAATTTTAAACAAACGACACAGTGCACT	1150
Qy	201	LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly	220
Db	1151	TTGTGTGTCTCTCCCCAGACATATGACATATATGACATTAATGATTTGACATGATGTG	1210
Qy	221	IleGluLysArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu	240
Db	1211	ATTAAAGACTACGCTGCTGTGGCATTTTACATGACATRGAAGGAGCATCCGCTATGAA	1270
Qy	241	ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal	260
Db	1271	CCAAAATGTCTCTTGGAGTACATCATCAAAAAGAGTTGGAGAAAGCTTCCATTAAGTA	1330
Qy	261	ProGlnLeuLysMetGluIle	267
Db	1331	CCCAATTTAAAAATGAGATA	1351

RESULT 4	
ABS76635	
ID	ABS76635 standard; DNA; 2197 BP.
XX	
AC	ABS76635;

XX	11-DEC-2002	(first entry)
DT		

DE DNA encoding novel human metalloprotease MPI.

KM Metalloproteinase, MP-1; immune disorder; glutamate transport; cancer;  
KM motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KM reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KM genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KM Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KM Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KM liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KM acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KM emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
KM neurological disorder; gene; ds.

**OS Homo sapiens.**

PN W0200272751-A2.

PD 19-SEP-2002

PF 05-FEB-2002; 2002WO-US003353.

PR. 05-FEB-2001; 2001US-0266518P.

PR 10-APR-2001; 2001US-0282814P.

PA (BRIM ) BRISTOL-MYERS SQUIBB CO

PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;

DR WPI; 2002-723329/78.

DR P-PSDB; ABG96478.

PT New isolated nucleic acid encoding MF-1 protein, useful for preventing, treating, or ameliorating diseases associated with aberrant metalloproteinase activity, e.g. immune, inflammatory and neurological disorders.

PS Claim 1; Fig 1A-C; 473pp; English.

The invention describes an isolated nucleic acid molecule (I) encoding a metalloprotease (Mpl-1). (I) is useful for preventing, treating, or ameliorating a medical condition, particularly an immune disorder, an aberrant glutamate transport or motor neuron disorder, such as amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like condition. The compositions and methods are also useful for diagnosing, prognosticating, treating, ameliorating and/or treating disorders associated with Mpl-1 activity, e.g. diabetes, cancer, reproductive disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome, or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease, Parkinson's disease, Huntington's disease or Tourette syndrome), liver and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic fibrosis) and vascular, inflammatory and neurological disorders (e.g. Alzheimer's disease or Parkinson's disease). This sequence represents a metalloprotease Mpl polynucleotide

Sequence 2197 BP; 681 A; 441 C; 439 G; 636 T; 0 U; 0 Other;

Alignment Scores:	
Pred. No.:	8.75e-147
Score:	1385.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	6
Gaps:	0
Length:	2197
Matches:	267
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-649-273-2 COPY 148 414 (1-267) x ABS76635 (1-2197)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheVal 20



QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
 DB 585 ATGAGGCTCAGTCACTACTATTAGTTGACCAATAAGTAATTCCTTTTAACTT 644  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
 DB 645 CTTTGTGATTTCTGGAGGTCACGTCTGTGGTCATTAGTTCAAGGAGTTTCAAGATTTTCTG 704  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
 DB 705 CTTCTGGAAGCTCTTTGGACATAGCACCGAGTGACATGCTTGACAGAGTGCGCAAGAA 764  
 QY 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80  
 DB 765 CTTTCTTTAATAAATCAATCCAGAGTCTCCACATGAGTGGTGGAAAGCCATTAGAACAT 824  
 QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
 DB 825 TTGGCCAAACAAAGAAATAGATTTCATTTTGACATCAACCTCCCTTGCATCATCTAAA 884  
 QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
 DB 885 AATTGTGATTTTCTTTACTGAGCTTCAACAGTTACTGATTAATAATATATGAAAG 944  
 QY 121 GluLysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
 DB 945 GAAAAAGGAGAAATATAGAGAGGCGAAATCCTGCTCTTCAGCAGCAGACATTCCTGCC 1004  
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
 DB 1005 ACAGTACAGCACACAATGCGCATGTCTGTGAAAAGAACACATCGGGCTATTCTGTTT 1064  
 QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValAlaLeuValAlaSerGlyGlyVal 180  
 DB 1065 TGTAGGACGAGAGACTGTGTACTCAAAATATATGAGTATGCTGTGTGTGTC 1124  
 QY 181 AlaSerAsnPheTYR11leArgArgAlaLeuGlnIleLeuThrAspAlaThrGlnCysThr 200  
 DB 1125 GCAAGTAACTTCTAATTCGCGAGAGCTCTGGAAATTTTACAAAGCCAAACAGATGCACT 1184  
 QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
 DB 1185 TTGTTGTCTCCCTCCAGACTATGACATGATATGAGCATATATGATGATGATGATGAT 1244  
 QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTYR12 240  
 DB 1245 ATGAAAAACATACGCTGGCTGTGGCATTTTACATGATGAAAGGCAATCCGCATATGAA 1304  
 QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
 DB 1305 CCAAAATGTCCTCTTGGAGTACATATCAAAAGAGTTGGAAAGCTTCCATTAAGATA 1364  
 QY 261 ProGlnLeuLysMetGluIle 267  
 DB 1365 CCACATATTAATAATGAGATA 1385

## RESULT 6

AAID46856  
 ID AAD46856 standard; cDNA, 1820 BP.

AC AAD46856;

DT 27-JAN-2003 (first entry)

DE Human glycoprotease 28472 cDNA.

KM Human; adenosine deaminase; seven transmembrane domain receptor; cancer;  
 KM 7TM; glycoprotease; immune disorder; Iga deficiency; allergy; arthyrhnia;  
 KM rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;  
 KM hyperextension; ischaemic heart disease; obesity; myocardial infarction;  
 KM endothelial cell disorder; Grave's disease; psoriasis; brain disorder;  
 KM Parkinson's disease; Alzheimer's disease; haematopoietic disorder;  
 KM cerebral oedema; metabolic disorder; liver disorder; platelet disorder;

KM chromosome mapping; tissue typing; gene therapy; neuroprotective;  
 KM cytosolic; anorectic; cardiac; haemostatic; gene; ss.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 FT CDS 146..1390  
 FT /tag= a  
 FT /product= "Human 28472 protein"  
 FT /note= "This region is specifically claimed as SEQ ID NO:  
 6 in claim 1 of the specification"

WO200274960-A2.

26-SEP-2002.

08-NOV-2001; 2001WO-US051427.

08-NOV-2000; 2000US-0246768P.

08-NOV-2000; 2000US-0246772P.

15-NOV-2000; 2000US-0249185P.

(MILL-) MILLENNium PHARM INC.

Leiby KR, Kapeller-Liebermann R, Glucksmann M;

WPI; 2002-759898/82.

P-PSDB; AAE29234.

New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,

useful for diagnosing and treating cancer, immune, cardiovascular,

hematopoietic, brain, pain, metabolic, liver or platelet disorders, and

in pharmacogenomics.

Claim 1; Fig 8; 178pp; English.

The present invention relates to novel 38650, 28472, 5495, 65507, 81588  
 or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-  
 protease or seven transmembrane domain (7TM) receptor family members.  
 Sequences of seven transmembrane domain (7TM) receptor family members.  
 or aberrant cellular proliferation and/or differentiation (e.g. colon or  
 lung cancer), immune disorders (e.g. selective Iga deficiency, rheumatoid  
 arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,  
 hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,  
 myocardial infarction, thrombus) including endothelial cell disorders  
 (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain  
 disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),  
 pain and metabolic disorders (e.g. obesity), liver disorders or platelet  
 disorders. They are also useful in screening assays, predictive medicine  
 (e.g. diagnostic assays, prognostic assays, monitoring clinical trials  
 and pharmacogenetics) and prophylactic and therapeutic methods. The  
 CC and nucleic acids may also be used in chromosome mapping, tissue typing and  
 CC forensic biology and as surrogate markers. Sequences of the invention are  
 CC also used in gene therapy. The present sequence is human glycoprotease  
 28472 cDNA

Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;

## Alignment Scores:

Score: 1.07e-142 Length: 1820  
 Percent Similarity: 1348.00 Matches: 260  
 Best Local Similarity: 98.508 Conservative: 3  
 Query Match: 97.388 Mismatches: 4  
 DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x AAD46856 (1-1820)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
 DB 585 ATGAGGCTCAGTCACTACTATTAGTTGACCAATAAGTAATTCCTTTTAACTT 644  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40

```
Db 647 CTTTGGATTTCTGAGGCTCACTGCTGTTGGCATATGTTCAAGAGGTTTCAGATTTTCG 706
Qy LeuLeuGlyYsSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaIArgArg 60
Db 707 CTTCTTGGAAAGCTTTTGGACATAGCACAGGTGACAGCTTGGACAGGTGCAAGAA 766
Qy LeuSerLeuIleYsHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
Db 767 CTTTCTTAATTAACATCCAGAGTGTCTCCACATGATGTGTGGAAAGCCATAGAACT 826
Qy LeuAlaIysGlnGlyAsnArgPheHisPheAspIleYsProProLeuHisHisAlaIys 100
Db 827 TTGGCCAAACAGAAATAGATTTCATTTGACATCAAACTCCCTTGATCATCTCTAA 886
Qy AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIysIleIleMetCysIys 120
Db 887 AATTGTGATTTTCTTTTACTGACCTTCACACCTTACTGATTAATAATGAAAAACAG 946
Qy GluIysGlnGluGlnIleGluIysGlnIleLeuSerSerAlaAlaAspIleAlaIa 140
Db 947 AAACAAAGAGAGGATTTGAGAGAGGCAATCTGTTTTCAGCAGCAGCATTTCTCC 1006
Qy TheValGlnHisThrMetAlaCysHisIleValIysArgThrHisArgAlaAlaIleuPhe 160
Db 1007 ACAGTACAGACACAATGGCATGTCTCTTGTGAATAAGAACACANTCGGGCTATTCTGTTT 1066
Qy CysIysGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyGlyVal 180
Db 1067 TGTAAAGCAGAGAGCTGTGTACCTCAAAATATGACATGCTGTGCTGTGCTGCTC 1126
Qy AlaSerAsnHisThrIleArgArgAlaIleuGlnIleLeuThrAsnAlaIleThrGlnCysThr 200
Db 1127 GCAAGTACTCTCTATATCCGAGAGCTCTGAAATTTTAAACAAAGCAACACAGTGCCT 1186
Qy LeuLeuCysProProProAlaGlyLeuCysThrAspAsnGlyIleMetIleAlaIArgAsnGly 220
Db 1187 TTGTGTGTCTCTCTCCACAGCTATGCACTGATATATGCACTTATGATTCATGAAATGT 1246
Qy IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
Db 1247 ATGAAAGACTACGCTGCTGGCTTGTGACATTTTACATGACATAGAAAGCATCCGCTATGAA 1306
Qy ProIysCysProLeuGlyValAlaAspIleSerIysGlyValGlyValAlaSerIleYsVal 260
Db 1307 CCAAAAGTCTCTCTTGGAGTAGACATATCAAAAGAGTTGAGAGAGCTTCCATTAAGTA 1366
Qy ProGlnLeuIysMetGluIle 267
Db 1367 CCACAATTAAATGAGATTA 1387

RESULT 7
ACA60887
ID ACA60887 standard; cDNA, 1820 BP.
XX
AC ACA60887;
XX
08-JUL-2003 (first entry)
XX
DE Human cDNA 28472 encoding a glycoprotease.
XX
KW Human; ss; gene; cancer; aberrant cellular proliferation;
KW differentiation; immune disorders; heart disorder; brain disorder;
KW cardiovascular disorder; endothelial cell disorder; pain disorder;
KW haematopoietic disorder; blood vessel disorder; metabolic disorder;
KW liver disorder; platelet disorder; glycoprotease.
OS Homo sapiens.
XX
XX key location/qualifiers
XX CDS 146..1390
XX FT /tag= a
XX FT /product= "Glycoprotease"
```

```
FT /note= "This CDS is specifically claimed in claim 1"
XX
PN US2003009017-A1.
XX
PD 09-JAN-2003.
XX
PF 08-NOV-2001; 2001US-00012140.
XX
PR 08-NOV-2000; 2000US-0246768P.
PR 08-NOV-2000; 2000US-0246772P.
PR 15-NOV-2000; 2000US-0249185P.
XX
PA (LEIB/) LEIBY K R.
PA (KAPE/) KAPELLER-LIBERMANN R.
PA (GLUC/) GLUCKSMANN M A.
PI
PI Leiby KR, Kapeller-Libermann R, Glucksmann MA;
PI WPI: 2003-428888/40.
DR P-PSDB; ABU09569.
XX
XX New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid
PT molecules, useful for diagnosing, treating cancer, pain, or immune,
PT heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic
PT and liver disorders.
XX
PS Claim 2; Fig 8; 90pp; English.
XX
XX The invention relates to an isolated 38650 (encoding adenosine
CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7
CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or
CC a sequence which is at least 60% identical to the six nucleic acids or
CC their open reading frames, fragments of at least 15 nucleotides,
CC naturally occurring variants, or a DNA insert of the plasmid deposited
CC with the American Type Culture Collection as Accession No. not defined in
CC the specification, which encodes the amino acid sequence). Also included
CC are a host cell containing the nucleic acids (used to produce the
CC proteins), the encoded proteins, an antibody that selectively binds to
CC the polypeptide, and identifying a compound that binds to/modulates the
CC activity of the polypeptide. The nucleic acid molecules, polypeptides and
CC methods are useful for diagnosing, treating cancer, aberrant cellular
CC proliferation and/or differentiation, immune disorders, heart disorders,
CC cardiovascular disorders including endothelial cell disorders,
CC haematopoietic disorders, blood vessel disorders, brain disorders, pain
CC and metabolic disorders, liver disorders and platelet disorders (many
CC examples of these disorders are given in the specification). The present
CC sequence is the Human cDNA 28472 encoding a glycoprotease
XX
SQ Sequence 1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 1.07e-142 Length: 1820
Score: 1348.00 Matches: 260
Percent Similarity: 98.50% Conservative: 3
Best Local Similarity: 97.38% Mismatches: 4
Query Match: 97.33% Indels: 0
Gaps: 0
DB: 8

US-10-649-273-2_COPY_148_414 (1-267) x ACA60887 (1-1820)
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIysValGluPheProPheLeuVal 20
Db 587 ATGGAGGCTCAATGCACTTACTATTTAGTGTGACCAATTAAGTAGAATTTCTTTTATGTT 646
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db 647 CTTTGAATTTCTGAGAGCTCACTGCTGTGTTGGCATATGTTCAAGAGTTTCAGATTTTCG 706
Qy 41 LeuLeuGlyYsSerLeuAspIleAlaIArgArg 60
Db 707 CTTCTTGGAAAGCTTTTGGACATAGCACAGGTGACAGCTTGGACAGGTGCAAGAA 766
Qy 61 LeuSerLeuIleYsHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
```

Dd		767	CTTTCCTTAATAAACATCCAGAGTGCTCCACCATGAGTGTGGAAAACCATAGAACAT	826
Oy		81	LeuAlaIysGInGIYAsnArgPheHisIspHeaspIIleYsProProLeuHiSHIsAlaIys	100
Dd		827	TTCGGCAACAAGAAMATAGATTTCATTTTAGCATCAAACCTCCTTGATCATGTCTAAA	886
Oy		101	AenCyAspAPheserPethThcGIYLeuGlnHISValIThrApIylSIIeIIemelLylals	120
Dd		887	AAATGTGATTTTTCTTTTACTGAGCTTCAACACGTTACTGTATATAAATATGAAAAACG	946
Oy		121	GluTySGInGIUGInGIYIleglulYsgIYGInIIleuSerSerAlaaIasPIIalala	140
Dd		947	AAACAAGAGGAAGGATTTAGAGAGGGGCAAAATCTGTCTTACGACGACATTCGTGCC	1006
Oy		141	ThrValGlnHIsthrmecAlCYeshIseuValIysArgTHrHIsargAlaIIleupe	160
Dd		1007	ACAGTACGACACACAATGGCATGTGCATCTTGAAAAAAGAACACATCGGGCATTCGT	1066
Oy		161	CySlySGInIagAspleuLeuProGlnasnbnAlaValIIeuValAIaseGIYGIYAl	180
Dd		1067	TGTAAAGCAGAGAGACTTGTATTACTCAAAATATAGCATGTGGCATTCGTGGTGTC	1126
Oy		181	AlaseAsnPhetYrIIeArgArGalenuGInIIleuTHrAsnaAlathrGlCYesthr	200
Dd		1127	GCAAGTAACTTCATATATCCGAGAGCTGTGAAATTTTAAACAAGCAACAGACAGCAT	1186
Oy		201	LeuLeuCysProProProIargLeuCYsthrAspangIYIlemelIeaIATPAsngIY	220
Dd		1187	TTGTTGTCTCCTCCTCCACGACTATGACCTATATATGSCATTTATGATTCAGTAATGT	1246
Oy		221	IlegIuarGLuearGalagIYleuGIYIIleuHIsaPIIlegIugIYIIaSYTYGIn	240
Dd		1247	ATTGAAACATCACGCTGGCTGGGCATTTTACATGACATAGAAAGGCATCCGCTATAA	1306
Oy		241	ProIysCysProLeuGIYValAspIleserIysGIUValIGIYgluaIaserIIleYsVal	260
Dd		1307	COAAATATCCTCTTGGAGTAGACATACAAAGAAAGTGGAGAAAGCTTCCATATAAGTA	1366
Oy		261	ProGlnLeuIysMetGInIIle 267	
Dd		1367	CCACATTTAAATAATGGAGATA 1387	
<b>RESULT 8</b>				
ABSS7020	ID	ABSS7020	standard; cDNA; 1821 BP.	
XX	AC	ABSS7020;		
XX	DT	30-JAN-2003	(first entry)	
DE	CDNA	encoding novel human glycoprotease 28472.		
XX				
KW		Cancer; aberrant cell proliferation; aberrant cell differentiation;		
KM		breast cancer; ovarian cancer; prostate cancer; colon cancer;		
KW		lung cancer; immune disorder; heart disorder; cardiovascular disorder;		
KV		endothelial disorder; hematopoietic disorder; blood vessel disorder;		
KW		brain disorder; pain; metabolic disorder; liver disorder; diabetes;		
KV		platelet disorder; carcinoma; sarcoma; leukaemia; Hodgkin's disease;		
KM		autoimmune disorder; hypertension; atherosclerosis; heart failure;		
KW		myocardial infarction; ischemic heart disease; Crohn's disease;		
KV		Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;		
KW		cerebral ischemia; peripheral neuropathy; Alzheimer's disease;		
XX		Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene; ss.		
OS	Homo sapiens.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	147..1391		
FT		/tag= a		
FT		/product= "Glycoprotease 28472"		
XX		/note= "Specifically claimed in claim 1"		

PN	W020027233-A2.
XX	
PD	03-OCT-2002.
XX	
PF	08-NOV-2001; 2001WO-US046724.
XX	
PR	08-NOV-2000; 2000US-0246768P.
PR	08-NOV-2000; 2000US-0246772P.
XX	
PR	15-NOV-2000; 2000US-0249185P.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Leibyl KR, Kapeller-Libermann R, Glucksmann M;
XX	
DR	WPI: 2003-029938/02.
XX	
DR	P-PSDB; ABG71162.
XX	
PT	New adenosine deaminase, glycoprotease and seven transmembrane domain
PT	nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
PT	81588 and 14354, useful for treating e.g. leukemia, Hodgkin's disease or
PT	hypertension.
XX	
PS	Claim 2; Fig 8A-B; 178pp; English.
XX	
CC	The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
CC	14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
CC	38650 nucleic acid molecule comprises a sequence encoding adenosine
CC	deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
CC	a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
CC	sequences that encode a human seven transmembrane domain (7TM). The
CC	38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
CC	sequences are useful for diagnosing, preventing or treating a subject
CC	with or at risk of developing a disorder, e.g. cancer or aberrant
CC	cellular proliferation and/or differentiation (e.g. breast, ovarian,
CC	prostate, colon or lung cancer), immune disorders, heart disorders,
CC	cardiovascular disorders, endothelial disorders, hematopoietic disorders,
CC	blood vessel disorders, brain disorders, pain and metabolic disorders,
CC	liver disorders or platelet disorders. These disorders include carcinoma,
CC	sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,
CC	hyperextension, atherosclerosis, heart failure, myocardial infarction,
CC	ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki
CC	syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral
CC	neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
CC	cachexia or diabetes. This sequence encodes the novel human glycoprotease
CC	28472
XX	
SO	Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;
Alignment Scores:	
Pred. No.:	1.07e-142 Length: 1821
Score:	1348.00 Matches: 260
Percent Similarity:	98.50% Conservative: 3
Best Local Similarity:	97.38% Mismatches: 4
Query Match:	97.33% Indels: 0
DB:	10 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x ABS57020 (1-1821)	
QY	1 MetGuaIAHAsAlaLeuThrIleArgLeuTTrpAsnIyValIGluPheProPheLeuVal 20
DB	588 ATGGAGGGCTCAAGCACTTACATTAAGTTGACCAATAAAGTAAAGAAATTTCTTTTAAATT 647
QY	21 LeuLeuIleSerGIgIYHISCySLeuLeuAlaLeuValIGInGIyValSerAsPheLeu 40
DB	648 CTTTGGATTTCTGGAGGTCACGTCTGTTGGCATTAAGTTCAGAGAGTTTTCAGATTTTCTG 707
QY	41 LeuLeuGIyLYeSerLeuAspIleAlaPArgIyAspMetLeuAspIyValAlaArgArg 60
DB	708 CTTCTTGAAGAAGTCTTTGGACATACACACAGGTGCATCTTGCACAAAGGTGGCAAGAGA 767
QY	61 LeuSerLeuIleLYeHISArgProGluCySerTrpMetSerGIyGIyValAlaIleGIuHIS 80
DB	768 CTTTCTTAAATTAACATTCGAAGAGGCTCACCATGAGGTGGTGGAAAGCCATPAGAACAT 827



QY 81 LeuAlaIySGInGIyAsnArgPheHisPheAspIleIySProProLeuHisHisAlaIyS 100  
 DB 828 TTGGCCAAAGAGAGAAATGATTTGATTCATCAACACCTCCCTTGATCATCTAA 887  
 QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValIThrAspLysIleIleMetLysLys 120  
 DB 888 AATGTGATTTTCTTTTACTGAGCTTCAACGCTTACTGATTAATAATGATAAAGG 947  
 QY 121 GlnIySGInGIyGlyIleGlnIySGInGIyIleLeuSerSerAlaIAspIleAla 140  
 DB 948 AAAAAGAGAGAGAGATTTGAGAGAGGCAAAATCTGCTTCAAGCAGACATTCCTGC 1007  
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIySArgThrHisArgAlaIleLeuPhe 160  
 DB 1008 ACAGTACAGCACACAAATGCAATGTCATCTTGTGAAAAGAACATCGGCTATTTGTTT 1067  
 QY 161 CysIySGInArgAspLeuLeuProGlnAsnAlaValIleuValAlaSerGlyGlyVal 180  
 DB 1068 TGTAAAGCAGAGAGACTTGTACTTCAAAATATGACAGTACTGTCATCTGCTGTC 1127  
 QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaIleThrGlnCysThr 200  
 DB 1128 GCAAGTAACTTCTATATCCGACAGCTCTGCAAAATTTTAACAAAGCAACAGTGCAC 1187  
 QY 201 LeuLeuCySPProProArgLeuCySPThrAspAsnGlyIleMetIleAlaIleThrAsnGly 220  
 DB 1188 TTGTTGTGCTCTCTCCAGACTATGACATGATTAATGCAATTAATGATGCAATGATGT 1247  
 QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrglu 240  
 DB 1248 ATTGAAGAGCTACGTCGTGGCTTGGCATTTTACATGACATGAAAGGATCCGCTATGA 1307  
 QY 241 ProIySCysProLeuGlyValAspIleSerIySGInGlyValIleGlyAlaSerIleLysVal 260  
 DB 1308 CCAAAAGTCTCTCTTGAGAGACATATCAAAAGAGTTGAGAGAGCTTCATAAAGTA 1367  
 QY 261 ProGlnLeuIyMetGluIle 267  
 DB 1368 CCACAATTAATAAATGAGATA 1388  
 RESULT 9  
 ADA52832  
 ID ADA52832 standard; cDNA; 2208 BP.  
 AC ADA52832;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Human coding sequence, SEQ ID 400.  
 XX  
 KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
 KW Gene Therapy; human; secretory protein; membrane proteins; cancer;  
 KW inflammatory disease; osteoporosis; neurological disease; gene; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 PN EP1293569-A2.  
 XX  
 PD 19-MAR-2003.  
 XX  
 PF 21-MAR-2002; 2002BP-00006586.  
 XX  
 PR 14-SEP-2001; 2001JP-00328381.  
 PR 24-JAN-2002; 2002US-0350435P.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PA  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 PI Isegai T, Sugiyama T, Otsuki F, Wakamatsu A, Sato H, Ishii S,  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Maeuho Y;  
 XX

DR WPI; 2003-395539/38.  
 DR P-PSDB; ADA544471.  
 XX  
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.  
 PS  
 PS Claim 1, SEQ ID NO 400; 205pp; English.  
 XX  
 CC The present invention relates to novel human secretory or membrane  
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.  
 XX  
 SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 3,3e-126 Length: 2208  
 Score: 1204.00 Matches: 239  
 Percent Similarity: 89.51% Conservative: 0  
 Best Local Similarity: 89.51% Mismatches: 4  
 Query Match: 86.93% Indels: 24  
 DB: 10 Gaps: 1  
 US-10-649-273-2\_COPY\_148\_414 (1-267) x ADA52832 (1-2208)  
 QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal 20  
 DB 785 ATGGAGGCTCAGACACTTATCTATTAGTTGACCAATAAAGTGAATTTCTTTTACGTT 844  
 QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
 DB 845 CTTTGTATTTCTGAGAGCTACTGCTGTTGGCATTAAGTTCACAGAGCTTTCAGATTTCTG 904  
 QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
 DB 905 CTTCTTGAAAGCTCTTTGGACATGACACAGGTGATCTTTCACAGAGTGGCAAGAA 964  
 QY 61 LeuSerLeuIleLysHisPheGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80  
 DB 965 CTTCTTTAATTAATCAATCAGAGTCTCCACATGATGATGAGGAAACCATTAAGACAT 1024  
 QY 81 LeuAlaIySGInGIyAsnArgPheHisPheAspIleIySProProLeuHisHisAlaIyS 100  
 DB 1025 TTGGCCAAAGAGAGAAATGATTTGATTCATCAACACCTCCCTTGATCATCTAA 1084  
 QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValIThrAspLysIleIleMetLysLys 120  
 DB 1085 AATGTGATTTTCTTTTACTGAGCTTCAACGCTTACTGATTAATAATGATAAAGG 1144  
 QY 121 GlnIySGInGIyGlyIleGlnIySGInGIyIleLeuSerSerAlaIAspIleAla 140  
 DB 1145 GAAAAGAGAGAGAGATTTGAGAGAGGCAAAATCTGCTTCAAGCAGACATTCCTGC 1204  
 QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIySArgThrHisArgAlaIleLeuPhe 160  
 DB 1205 ACAGTACAGCACACAAATGCAATGTCATCTTGTGAAAAGAACATCGGCTATTCGTT 1264  
 QY 161 CysIySGInArgAspLeuLeuProGlnAsnAlaValIleuValAlaSerGlyGlyVal 180  
 DB 1265 TGTAAAGCAGAGAGACTTGTACTTCAAAATATGACAGTACTGTCATCTGCTGTC 1324  
 QY 181 AlaSerAsnPheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaIleThrGlnCysThr 200  
 DB 1325 GCAAGTAACTTCTATATCCGACAGCTCTGCAAAATTTTAACAAAGCAACAGTGCAC 1384  
 QY 201 LeuLeuCySPProProArgLeuCySPThrAspAsnGlyIleMetIleAlaIleThrAsnGly 220  
 DB 1385 TTGTTGTGCTCTCTCCAGACTATGACATGATTAATGCAATTAATGATGATCA----- 1435  
 QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrglu 240





XX Homo sapiens.  
OS  
PN W02003062376-A2.  
XX  
PD 31-JUL-2003.  
XX  
PF 13-JAN-2003; 2003WO-US001096.  
XX  
PR 16-JAN-2002; 2002US-0349384P.  
PR 17-JAN-2002; 2002US-0349413P.  
PR 17-JAN-2002; 2002US-034946P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JL,  
PI Yu YJ, Thason O, Yap PE, Amshay SR, Dam TC, Liu TF, Gerstin EH,  
PI Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RX, Urashka MB,  
PI Kristnam SR, Kolluru V, Panesar IS;  
XX  
XX WPI: 2003-636732/60.  
DR P-PSDB; ADE31156.  
XX  
PT New human diagnostic and therapeutic polynucleotides and polypeptides,  
PT useful for diagnosing, treating or preventing e.g. leukemia, brain  
PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke  
PT or Alzheimer's.  
XX  
PS Claim 1; SEQ ID NO 100; 634bp; English.  
XX  
XX The invention relates to a novel isolated human diagnostic and  
XX therapeutic polynucleotide (designated ditbp). The novel ditbp  
XX polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798  
XX base pairs fully defined in the specification; a polynucleotide  
XX comprising a naturally occurring polynucleotide sequence at least 90%  
XX identical to the ditbp polynucleotide; a polynucleotide complementary to  
XX the ditbp polynucleotide or its polynucleotide which is at least 90%  
XX identical; or an RNA equivalent of any of the polynucleotides mentioned  
XX above. The ditbp polynucleotides have the following activities:  
XX antileptospiretic, antiinflammatory, cerebroprotective, antilipemic,  
XX antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,  
XX tranquilizer, osteopathic, antiarthritic, antineumatic, cyostatic,  
XX hepatocytic, virucide, haemostatic, anti-HIV, antithyroid, thymomimetic,  
XX chemoprotective, antibacterial, fungicide, antiparasitic, anticonvulsant,  
XX thrombolytic, anticoagulant, anorectic, vasotropic, and anticancer. The  
XX novel DITBP polynucleotides polypeptide can be used in gene therapy and  
XX protein replacement therapy. The ditbp polynucleotides or DITBP  
XX polypeptides are useful for diagnosing, preventing or treating diseases  
XX associated with the expression of human molecules. In particular, these  
XX diseases include cancers (e.g. adenocarcinoma, leukaemia, melanoma, brain  
XX cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung  
XX cancer) or other cell proliferative disorders (e.g. arteriosclerosis,  
XX atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary  
XX thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,  
XX Addison's disease, thyroiditis, Crohn's disease, Graves' disease,  
XX Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid  
XX arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.  
XX viral, bacterial, fungal or parasitic infection), developmental disorders  
XX (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.  
XX thrombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic  
XX disorders (e.g. hypercholesterolemia, hypoglycaemia, diabetes,  
XX hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic  
XX cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,  
XX Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,  
XX anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),  
XX transport disorders (e.g. akinesia or multidrug resistance), or  
XX connective tissue disorders (e.g. Paget's disease or rickets). This  
XX polynucleotide sequence represents one of the human ditbp DNA sequences  
XX of the invention.  
SQ Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;

Pred. No.: 2,99e-102 Length: 3358  
Score: 995.50 Matches: 250  
Percent Similarity: 44.82% Conservative: 1  
Best Local Similarity: 44.64% Mismatches: 7  
Query Match: 71.88% Indels: 307  
DB: 10 Gaps: 2  
US-10-649-273-2\_COPY\_148\_414 (1-267) x ADE31345 (1-3358)  
QY 10 LeuThraenlyValIGlnpheProPheLeuValLeuLeuIIseerGlyGlyHisCysLeu 29  
DB 3067 GTGACCAATTAAGTGAATTCCTTTTATGTTCT-TTGATTTCTGAGGTCACTGCTG 3009  
QY 30 LeuAlaleuValIGlYValSerAppPheLeuLeuGlyYHisSerLeuAspIleAla 49  
DB 3008 TGGGCAATTGTTCAAGAGTTTCAGATTTCCTTCTTGGAAGTTC-TTGAACATAGCA 2950  
QY 50 ProGlyAspMet----- 53  
DB 2949 CCAAGTGACATCGTTGACAGGTAATTAAAGATTTCCTTCATTCTTTTGTATGT 2890  
QY 53 ----- 53  
DB 2889 TGTCCATTTCACCTAAGTGAATGATGCTACCACTTCACCTAATAATTCTGAA 2830  
QY 53 ----- 53  
DB 2829 TTTTATCTTAGTAACCTGAACAAANAATTCACATATGTGAGAAAAAATAGAACAGTA 2770  
QY 53 ----- 53  
DB 2769 GTACAGCAATTATTAATTTCTTAGCCTTTCTTAATAAATGTAAGAGTTCAATCTGTA 2710  
QY 53 ----- 53  
DB 2709 CATAAAGCTGAATAGTTTGACAGATACAGTTATGTATTTGCCAANAATAGTATGTA 2650  
QY 53 ----- 53  
DB 2649 AGAAGCTGCTGTAACTAACAATGCAAAAAAGTAAATAGAGAAATATATATAGA 2590  
QY 53 ----- 53  
DB 2589 TTAACATAAGACATTAAGAATGCAATGACAGAAATTAATGACAAATTACTTACACA 2530  
QY 53 ----- 53  
DB 2529 CAGACAGGTCCTCCCGACCCCTTTGTTAGAACTACAGAGGCTACTGCATAT 2470  
QY 53 ----- 53  
DB 2469 ATAGAGAACTACAGACAGACAGACAGTCTTCCACAGTGAATATAGAAATATAGG 2410  
QY 53 ----- 53  
DB 2409 ACAAGTTCTTATTTATGACGTTCAATCATTAAGACGTTATGCACTTCAAGCCATT 2350  
QY 53 ----- 53  
DB 2349 CCAACCAATAGAGAGCAAAACATAGACAGGGGAGATGTTGGCTTATTTGTTGGGTC 2290  
QY 53 ----- 53  
DB 2289 ATCATAGGAACAGGGTGTCTGCTTACCTGAATATCACCTATATGCTATATTTGCCAA 2230  
QY 53 ----- 53  
DB 2229 GTATAGCAATTTTATTCATTCAGGGGTTTTTTGTTGTTAGTAATTTCAATTAAT 2170  
QY 54 -----Leu 54  
DB 2169 CCTTGACATTTTGTGTTGACAGTATTTAATTTATGACTGAATAAATATGTTCTTT 2110

```

QY 55 AsplysVal1a1aArgArgLeuSerLeu11e1y5h1sProGluCyseSerThrmSerGly 74
DB 2109 GATAG-GTGGCAAGAAGACTTCTTAATATAAACAATCCAGAGTGTCCACCAATGAGTGT 2051
QY 75 -GlyVala1a11eGlu1h1sLeuAla1yG1nG1yAsnArgPheh1sPheAsp11e1yPr 94
DB 2050 GGGGAAAGCCATAGAACATTTGGCCAAACAGAAATATGATTTCTTTGACATCAAAAC 1991
QY 94 OProLeu1h1s1a1yAsnCyseAspPheSerPheThrg1yLeuGlu1h1sVal1Thrs 114
DB 1990 TCCCTGCAATCAATGCTAAATAATTTGATTTCTTTTACTGACCTTCAACAGTTACTGA 1931
QY 114 pLys11e1eMet1y5G1u1yG1u1yG1u1yG1u1yG1u1yG1u1yG1u1yG1u1yG1u1y 124
DB 1930 TAAATATATATGAAGAAAGAAAGAGAGATATTTCTTAATTAAGTAAGTTGAACA 1871
QY 125 -----Gly11eGlu1yG 129
DB 1870 GATAAATATTCCTGATGCTGCTTAAATAATAGCTGCTCATTTCTGACAGTTATGAGAAAG 1811
QY 129 LyG1u11eLeuSerSer1a1a1aAsp11e1a1a1aThrVal1G1n1h1sThrm1a1aCyseH 149
DB 1810 GGCATATCTGTTCTTCAAGACAGACATTTGCTGCAACATAGACACACAAATGGCAATGTC 1751
QY 149 1sLeuVal1yAsrThrh1sArgAla11eLeuPheCyse1yG1nArgAsp1eLeuProG 169
DB 1750 ATCTTGTAAGAAAGAACACATCGGCTATCTCTTTTGTAGACAGAGACTTGTATACCTC 1691
QY 169 1nAsnAsnAla1Val1eVal1a1a1aSerG1y1Val1a1a1aSerAsnPhery11eArgArg 189
DB 1690 AAAATAAGCAGTCTGCTGTCATCTGT- GGTGTGCAAGTAATCTTATATCCACAGAG 1632
QY 189 1a1eAsnGlu11eLeuThrmAsn1a1aThrg1nCyseThrmLeuCyseProP-PCAArgLeu 209
DB 1631 CTCTGGAATTTTAAAC-AAACGAAACAAGTGCATTTGTTGTCCTCTCCCAAGATAT 1573
QY 209 ysrThrAspAsnG1y11eMet11eAla1a1aTrpAsnG1y11eGluArgLeuArgAlaG1yLeuG 229
DB 1572 GCACGTGAATTAAGGCAATTAATGATGATGCAATGAATGATTAAGAACTACCTGCTGCTTG 1513
QY 229 1y11eLeuH1sAsp11eGluG1y11eArg1y11eArg1y11eArg1y11eArg1y11eArg1y 249
DB 1512 GCATTTTACATGACATAGACAGCATCCGCTATGAACCAAAATGCTCTTGTGAGTAGACA 1453
QY 249 1eSer1yG1u1a1G1yG1u1a1aSer11e1yVal1ProG1nLeu1yMetG1u1e 267
DB 1452 TATCAAAAAGAGTGGAGAGCTTCCATTAAGTACCAATTAATAAATGAGATA 1397

```

```

RESULT 12
ID ABQ75508 standard; DNA; 1572 BP.
AC ABQ75508;
DT 07-NOV-2002 (first entry)
XX
DE Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.
XX
KM Murine, mouse; protease; calcium activated neutral protease type 5;
KW CAPN5; trypsinase 4; sialoglycoprotease; enzyme; genetic disease;
KW neurological; neuropsychological; psychotic illness; transgenic animal;
KW gene; ds.
OS Mus musculus.
PN M0200245491-A2.
PD 13-JUN-2002.
PF 05-DEC-2001; 2001WO-USO46405.
PR 06-DEC-2000; 2000US-0251803P.
PR 06-DEC-2000; 2000US-0251820P.

```

```

PR 13-DEC-2000; 2000US-0255971P.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Allen KD, Leviten KM;
XX
DR WPI; 2002-657389/70.
XX
PT Novel transgenic animal, comprising a disruption in protease target gene,
PT is useful for identifying agents that ameliorates a phenotype associated
PT with a disruption in a protease target gene.
XX
PS Example 3; Fig 7; 62pp; English.
XX
CC The present invention describes a non-human transgenic animal (I)
CC comprising a disruption in a protease target gene (PG) selected from
CC calcium activated neutral protease type 5 (CAPN5) gene, trypsinase 4 gene
CC and sialoglycoprotease-like gene. Also described is a targeting construct
CC (II), comprising a first polynucleotide sequence homologous to at least a
CC first portion of PG, a second polynucleotide sequence homologous to at
CC least a second portion of PG and a selectable marker. (II) is useful for
CC producing a transgenic mouse comprising a disruption in a protease target
CC gene. By introducing (II) into a cell, introducing the cell into a
CC blastocyst, implanting the resulting blastocyst into a pseudopregnant
CC mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
CC and breeding the chimeric mouse to produce the transgenic mouse. (I) is
CC useful for identifying an agent that modulates the expression or function
CC of a protease target gene, by administering an agent to (I) and
CC determining whether the expression or function of the disrupted protease
CC target gene in (I) is modulated. (I) is also useful for testing the
CC efficacy of proposed genetic and pharmacological therapies for human
CC genetic diseases, such as neurological, neuropsychological or psychotic
CC illness. The present sequence represents murine sialoglycoprotease-like
CC gene sequence, which is used in an example from the present invention
XX
SQ Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;

```

```

Alignment Scores:
Pred. No.: 2,67e-101 Length: 1572
Score: 963.00 Matches: 210
Percent Similarity: 84.21% Conservative: 14
Best Local Similarity: 78.95% Mismatches: 38
Query Match: 70.97% Indels: 6
DB: 6 Gaps: 2

```

```

US-10-649-273-2_COPY_148_414 (1-267) x ABQ75508 (1-1572)
QY 1 MetGluAlaH1a1a1eLeuThr11eArgLeuThrAsn1yVal1G1uPheProPheLeuVal 20
DB 445 ATGAGAGCTCAAGCACTAGCTATTAAGCTCAACCAATTAAGTATGATTN-CTTTTATAGTT 503
QY 21 LeuLeu11eSerG1y1yH1sCyseLeuLeuAlaLeuVal1G1nG1yVal1SerAspPheLeu 40
DB 504 CTTTGTGATTTCTGGCGGACATGCTGCTGTGAKRTTAGTCCAAAGTGTTCCTCCATTTCCG 563
QY 41 LeuLeuG1y1ySer1eLeuAsp11eAla1a1aProG1yAspMetLeuAsp1yVal1a1aArgArg 60
DB 564 CTCTTGGAGATCTTTTGACATAGACACAGGCGACATCTTGACAAAGGTGGCAAAAGA 623
QY 61 LeuSer1eLeu11e1yH1sProG1uCyseSerThrmSerG1y1yVala1a11eGluH1s 80
DB 624 CTTTCTTTAATCAACATCCAGAAATGTTTCAATGAGTGTGGGAAAGACTTATGAAACG 683
QY 81 LeuAla1yG1nG1yAsnArgPheh1sPheAsp11e1yProProLeuH1s1a1a1yS 100
DB 684 TTGGCCAAAGAGCAATATGATTTCAATTTTACATCAATCCATTAATGCAATGCTAAG 743
QY 101 AsnCyseAspPheSerPheThrg1yLeuGlu1h1sVal1ThrsPly11e1eMet1y51y 120
DB 744 AATTCGATTTTCTTTCAACGAGACTTCAACATATTTACTGATAGCTAATTAACACAAG 803
QY 121 Glu1yG1u1GluG1y11eGlu1yG1nG11eLeuSerSerAla1a1aAsp11eAla1a 140

```

804 GAAAAAGAGAGCATTTGAGAGAGGCGCAATCTGTGATCATGCTGACAGATTCTGCT 863  
QY 141 ThValGlnHisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleuPhe 160  
Db 864 GCGGTACAGCATGACAGACGCTGCACCTTGCAGAAAGAACATCGCGCTATTCTGTTT 923  
QY 161 CysLeuGlnArgAspLeuLeuPheGlnAsnAsnAlaValIleuValAlaSerGlyVal 180  
Db 924 TGCAGAGCAAAAATTTGCTCTCTCCAGCTAACCCAGATTAAGTTGATCTCGAGTGT 983  
QY 181 AlSerAsnPheryIleArgArgAlaLeuGlnIleuThrAsnAlaThrGlnCysThr 200  
Db 984 GCAAGTAATCTGTATCATCCGAAAGCATTTGGAATTTGCGCAATGCAACGCATGACGG 1043  
QY 201 LeuLeuCysProPheProAlaArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
Db 1044 TTGTTGTCTCACT-TCAGACTGTGCACTGACAAATGGCATATGATTCAGATGAGA 1102  
QY 221 TleGluArgLeuArgAlaGlyLeuGlyIleLeu-HisAsp-TleGluGlyIleArgTyrG 240  
Db 1103 ATTGGAAGATTACGTGCTGCTGCTTGGCTTTACCATATGATAGAGACATTCGGTTATT 1162  
QY 240 IuProIysCysProLeuGlyVal--AspIleSerIysGluValGlyIuIaSerIleL 259  
Db 1163 AACCAATATGCTCTTGTGAGTGAAGGATATATCCGAAAGAGTTGGCAGA---AGCTTGC 1219  
QY 259 ysValPheGlnLeu 263  
Db 1220 CCATTAAAAAGTTA 1233

## RESULT 13

AAS84622  
ID AAS84622 standard; cDNA; 2734 BP.

AC AAS84622;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #20426.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

XX MO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001W0-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG20435.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity.

XX Claim 1, SEQ ID NO 20426, 103bp, English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX sequences, (i) is useful as hybridisation probes, polymerase chain

XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX and in recombinant production of (II). The polynucleotides are also used

XX in diagnostics as expressed sequence tags for identifying expressed

CC genes, (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 2734 BP; 773 A; 545 C; 639 G; 763 T; 0 U; 14 Other;

## Alignment Scores:

Pred. No.:	2,61e-57	Length:	2734
Score:	599.00	Matches:	131
Percent Similarity:	71.57%	Conservative:	10
Best Local Similarity:	66.50%	Mismatches:	22
Query Match:	43.25%	Indels:	34
DB:	5	Gaps:	6

US-10-649-273-2\_COPY\_148\_414 (1-267) x AAS84622 (1-2734)

QY 71 ThrMetSerGlyGlyLeuAlaIleGluHisLeuAlaIleGluGlyAsnArgPheHisPhe 90

Db 206 ACCCTGCTAAGGGGGGCGCCCT-----AGATTCCAT----- 235

QY 91 AspIleLysProProLeuHisHisAlaLys-----AsnCys----- 102

Db 236 -----CCTATTCTCCGATGAAGATGACGCTGTGTAGAACGTGCGACGA 283

QY 103 -----AspPheSerPheThrGlyLeuGlnHisValThrAsp 114

Db 284 TCTGCTTCGGAGAGCCCGAGAGAGCTTTTACTCATGCGGAGAAAGATGGAGCTAGG 343

QY 115 LysIleIleMetLysLeuGluLysGluGlu-----GlyTleGluLysGlnIleLeu 132

Db 344 GCTTTACATGCGACAGAGACGACCGAGAGAGGGGGGATTTGAGAGAGGGGCAATCTCG 403

QY 133 SerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLys 152

Db 404 TCTTCAGACAGACATGCTGCCACAGTACAGCACACAAATGTCATCTTGTGAAA 463

QY 153 ArgThrHisArgAlaIleLeuPheCysLysGlnAspLeuLeuProGlnAsnAsnAla 172

Db 464 AGAACACATCGGGCTATTCTGTTTGTAAAGAGAGACTTGTACTCAAAATATATGCA 523

QY 173 ValLeuValAlaSerGlyGlyValAlaSerAsnPheryIleArgArgAlaLeuGluIle 192

Db 524 GTACTGTTGCACTGCGTGTGTCGAAAGTAACTTATATCGGACGCTCGGAAATT 583

QY 193 LeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsn 212

Db 584 TTACCAACGCAACACAGTGCCTTGTGTCTCTCCAGACTATGCACTGTAAT 643

QY 213 GlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleLeuHis 232

Db 644 GGCATTATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 703

QY 233 AspIleGluGlyIleArgTyrGluProLys-----CysProLeuGlyVal 247

Db 704 GACATGAAGAGCATCGCTATGACCAAGATATGTCCTCTTACGGGCTG 754

## RESULT 14

ADL86725  
ID ADL86725 standard; DNA; 371 BP.

XX

AC	ADL86725;
XX	
DT	20-MAY-2004 (first entry)
XX	
DE	DNA up-regulated in murine multipotent progenitor cells SeqID 3118.
XX	
KW	gene potential; multi-lineage; cell commitment; haematopoietic stem cell;
KM	HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;
KX	common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.
XX	
OS	Mus sp.
XX	
PN	WO2003093445-A2.
XX	
PD	13-NOV-2003.
XX	
PF	05-MAY-2003; 2003WO-US014114.
XX	
PR	03-MAY-2002; 2002US-0377383P.
XX	
PA	(STOW-) STOWERS INST MEDICAL RES.
XX	
P1	Li L;
XX	
DR	WPI; 2004-022656/02.
XX	
PT	Classifying an unknown multi-lineage affiliated gene comprises isolating
XX	expressed nucleic acid sequences from the discrete cell sub-populations.
PS	
XX	Claim 8; SEQ ID NO 3118; 123jp; English.
CC	This invention relates to a novel method for predicting gene potential by
CC	associating nucleic acid sequences of unknown function with particular
CC	sub-population profiles. Specifically, it refers to classifying an
CC	unknown multi-lineage affiliated gene by collecting hybridisation data to
CC	develop a gene expression map, in order to determine the discrete sub-
CC	population where it is expressed. The present invention describes methods
CC	for predicting the lineage commitment of genes associated with the self-
CC	renewing haematopoietic (blood) stem cells (HSCs), as well as the non-
CC	self renewing multipotent progenitors (MPPs), common lymphoid progenitors
CC	(CLPs) and common myeloid progenitors (CMPs), which are collectively
CC	referred to as bone marrow stem cells populations. As such, these methods
CC	can be used to identify associated multi-lineage affiliated genes and
CC	hence the underlying molecular mechanisms in physiological haematopoietic
CC	development. This polynucleotide sequence is DNA associated with a murine
CC	MPP sub population of cells of the invention.
XX	
SQ	Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;
Alignment Scores:	
Pred. No.:	1.18e-43 Length: 371
Score:	468.00 Matches: 92
Percent Similarity:	78.86% Conservatve: 5
Best Local Similarity:	74.80% Mismatches: 26
Query Match:	33.79% Indels: 0
DB:	12 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x ADL86725 (1-371)	
OY	137 AAPLEAIAA1aThrValGlnHisThrMetAlaCysHisLeuValLysArgTrpHisArg 156
Db	2 GACATTGGCTGCGGTGACGATGCATGCAGCGGTGCCACTTGCGAAAGAACAACATCGC 61
OY	157 AlalieleuPheCysIleGlnArgAspleuLeuProGlnAsnAsnAlaValleuValAla 176
Db	62 GCTATTCTGTTTGGACAGCAAGAAATTGGCTCTCCAGCTTAACCAAGTATTAGTTGA 121
OY	177 SerGIgLYValAlaSerAsnPheTYrIIeaRgaGAlaLeuGluIleLeuThrAsnAla 196
Db	122 ICTGGAGGGTGTGCAGATTAACCTGTACATCCAGAAAGCATTTGGAAATTCGCCAAATGCA 181
OY	197 ThrGInCYthrLeuCYsPrOProARgLeuCYthrAspAnGlylleMetIle 216

Db	182	AAGCAGTACAGTGTGTTGTGTCACCTCCAAACACTGTGACATGACAAATGGCATCANGATT	241
Qy	217	AlATPaPanglYlIEgluaRgleuAgaIyleuGIYlEdeuHIsapllIEgluGIy	236
Db	242	GCATGGAATGGAATTAAGANGATTACGTGCGNGCTTNGGCTNTTTCATCMTGANNATGAAAGAC	301
Qy	237	IlEhAgTYRgIuProlYsCySProlEuglYVaIsapllIEserlysgIuVaIGIygluAa	256
Db	302	ATCCNNATATGNACCAAAAATTCNTNTTGTGAGTAGACATNTCCAGAAAGTGTGCAAGAGCT	361
Qy	257	SerIIElyS	259
Db	362	GCACATAAA	370
RESULT 15			
ID	ADL86726	ADL86726 standard; DNA; 371 BP.	
AC	ADL86726;		
XX			
DT	20-MAY-2004	(first entry)	
DE		DNA up-regulated in murine multipotent progenitor cells SegID 3119.	
XX			
KW		gene potential; multi-lineage; cell commitment; haematopoietic stem cell;	
KW		HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;	
OS		common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.	
XX			
XX	Mus sp.		
PN			
XX	W02003093445-A2.		
PD			
XX	13-NOV-2003.		
PF			
XX	05-MAY-2003; 2003MO-US014114.		
PR			
XX	03-MAY-2002; 2002US-0377383P.		
PA			
XX	(STOW-) STOWERS INST MEDICAL RES.		
PI			
XX	L1 L;		
DR			
XX	WPI; 2004-022656/02.		
XX			
XX			
PS		Claim 8; SEQ ID NO 3119; 123pp; English.	
XX			
CC		This invention relates to a novel method for predicting gene potential by	
CC		associating nucleic acid sequences of unknown function with particular	
CC		sub-population profiles. Specifically, it refers to classifying an	
CC		unknown multi-lineage affiliated gene by collecting hybridisation data to	
CC		develop a gene expression map, in order to determine the discrete sub-	
CC		population where it is expressed. The present invention describes methods	
CC		for predicting the lineage commitment of genes associated with the self-	
CC		renewing haematopoietic (blood) stem cells (HSCs), as well as the non-	
CC		self renewing multipotent progenitors (MPPs), common lymphoid progenitors	
CC		(CLPs) and common myeloid progenitors (CMPs), which are collectively	
CC		referred to as bone marrow stem cells populations. As such, these methods	
CC		can be used to identify associated multi-lineage affiliated genes and	
CC		hence the underlying molecular mechanisms in physiological hematopoietic	
CC		development. This polynucleotide sequence is DNA associated with a murine	
CC		MPP sub population of cells of the invention.	
XX			
XX			
SEQ		Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;	
Alignment Scores:			
Pred. NO.:	1.18e-43	Length:	371
Score:	468.00	Matches:	92
Percent Similarity:	78.86%	Conservative:	5
Best Local Similarity:	74.80%	Mismatches:	26
Query Match:	33.79%	Indels:	0

```

DB: 12 Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x ADJ86726 (1-371)
QY 137 Aspl1eAla1aThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArg 156
Db 2 GACATTCCTGCTGCGGTACAGCATGCACAGCGTCCACCTTGCGAAAAAGACACATCGC 61
QY 157 Ala1leuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAla 176
Db 62 GCTATTCCTGTTGCAAGCAGAGAAATTGCTCTCTCCAGCTAACGCAATATTAGTTGTA 121
QY 177 SerGlyGlyValAlaSerAsnPheTyr1leArgAlaLeuGln1leuThrAsnAla 196
Db 122 TCTGGAAGTGTGTCAGATTACTGTACATCCGAAAGCATTTGGAATTTGTCGCAAAATGCA 181
QY 197 ThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGly1leMet1le 216
Db 182 ACGCAGTGCACGTTGTTGTGTCCACCTCCAAAGACTGTGCACTGACATGCAATGCAATGATT 241
QY 217 AlaThrAsnGly1leGlnArgLeuArgAlaGlyLeuGly1leLeuHisAsp1leGlnGly 236
Db 242 GCATGGAATGGAATTGAANGATTACGTGCNGNCTTNGCNTTNTACNTGANNNTAGAGAC 301
QY 237 1leArgTyrGlnProLysCysProLeuGlyValAsp1leSerLysGlyValGlyGlnAla 256
Db 302 ATCCNNTATGNAACCAAAATNTCNTTTTGAAGTAGACATNTCCAGAAAGTTGCAGAAAGCT 361
QY 257 Ser1leLys 259
Db 362 GCCATAAAA 370

```

Search completed: February 16, 2005, 13:58:48  
 Job time : 525.461 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:19:37 ; Search time 158.749 Seconds  
(without alignments)  
2752.056 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHATRLTNKVEFPFLV.....DISKVEGASIKVPLQKMEI 267

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 segs, 81818359 residues  
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+g2n.model-DEV=x1h  
-Q=/cgn2\_1/USPRO.epool/US10649273/runat\_14022005\_114704\_16425/app\_query.fasta\_1.1429  
-DB=Issued Patents NA -QEMT=fastcap -SUFFIX=rn1 -MINMATCH=0.1 -LOOFC=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USBR=US10649273 @CGN\_1\_1\_292 @runat\_14022005\_114704\_16425 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	1416	4	US-09-774-528-177 Sequence 177, App
2	1385	100.0	1526	4	US-10-067-443-23 Sequence 23, Appl
3	1385	100.0	2197	4	US-10-067-443-1 Sequence 1, Appl
4	1385	98.1	1387	4	US-10-067-443-21 Sequence 21, Appl
5	824.5	59.5	14364	4	US-10-067-443-20 Sequence 38, Appl
6	308.5	22.3	94750	4	US-09-596-002-38 Sequence 806, App
7	307	22.2	1053	4	US-09-540-236-806 Sequence 884, App
8	291.5	21.0	1059	4	US-09-252-991A-884 Sequence 801, Appl
9	291.5	21.0	1206	4	US-09-252-991A-801 Sequence 1, Appl
10	289.5	20.9	1830121	4	US-09-557-884-1 Sequence 1, Appl
11	289.5	20.9	1830121	4	US-09-643-990A-1 Sequence 6612, Ap
12	282	20.4	996	4	US-09-902-540-6612

C 13	282	20.4	2582	4	US-09-902-540-503	Sequence 503, App
14	280.5	20.3	1074	4	US-09-543-681A-2341	Sequence 2341, Ap
15	270.5	19.5	1315	1	US-08-087-797-1	Sequence 1, Appl
16	268.5	19.4	1032	4	US-09-489-039A-2050	Sequence 2050, Ap
17	261	18.8	1092	4	US-09-107-532A-2955	Sequence 2955, Ap
18	254	18.3	1008	3	US-08-987-121A-5	Sequence 5, Appl
19	254	18.3	1011	3	US-08-987-121A-3	Sequence 3, Appl
20	250	18.1	1006	3	US-08-961-083-51	Sequence 51, Appl
21	250	18.1	1006	4	US-09-536-784-51	Sequence 51, Appl
22	250	18.1	10974	3	US-08-961-527-214	Sequence 214, App
23	248	17.9	1011	3	US-09-066-512-1	Sequence 1, Appl
24	246	17.8	1011	4	US-09-583-110-2196	Sequence 2196, Ap
25	241	17.4	1011	4	US-09-107-433-1618	Sequence 1618, Ap
26	236.5	17.1	640681	4	US-09-790-988-1	Sequence 1, Appl
27	236	17.0	822	4	US-09-107-279-727	Sequence 727, App
28	236	17.0	1101	4	US-09-134-000C-1551	Sequence 1551, Ap
29	236	17.0	3993	4	US-09-110-279-3985	Sequence 3985, Ap
30	233	16.8	1107	3	US-09-134-001C-1072	Sequence 1072, Ap
31	220	15.9	1026	3	US-09-149-624-1	Sequence 1, Appl
32	218	15.7	3064	3	US-09-221-017B-794	Sequence 794, App
33	218	15.7	1230025	4	US-09-198-452A-1	Sequence 1, Appl
34	218	15.7	1230230	4	US-09-438-185A-1	Sequence 1, Appl
35	216	15.6	15249	4	US-08-956-171E-102	Sequence 102, App
36	216	15.6	15249	4	US-08-781-986A-102	Sequence 102, App
37	215.5	15.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl
38	215.5	15.6	4411529	3	US-09-103-840A-1	Sequence 147, App
39	211.5	15.3	11555	4	US-09-602-777A-147	Sequence 1, Appl
40	207.5	15.0	1664976	4	US-08-916-421B-1	Sequence 1, Appl
41	207.5	15.0	1664976	4	US-09-692-570-1	Sequence 6, Appl
42	192.5	13.9	1663	4	US-09-620-312D-6	Sequence 131, App
43	185.5	13.4	42325	4	US-08-311-731A-131	Sequence 261, App
44	181.5	13.1	432	4	US-09-328-352-261	Sequence 19, Appl
45	169	12.2	876	4	US-09-724-623-19	

ALIGNMENTS

RESULT 1  
US-09-774-528-177  
Sequence 177, Application US/09774528  
Patent No. 6743619  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Zhou, Ping  
APPLICANT: Goodrich, Ryle  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Ren, Feiyan  
APPLICANT: Zhang, Jie  
APPLICANT: Zhao, Qing A.  
APPLICANT: Yang, Yonghong  
APPLICANT: Xue, Aidong J.  
APPLICANT: Wehrman, Tom  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Wang, Dunrui  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. 6743619el Nucleic Acids and  
FILE REFERENCE: Polypeptides  
CURRENT PILING DATE: 2001-01-30  
NUMBER OF SEQ ID NOS: 441  
SOFTWARE: pt\_fl\_genes Version 2.0  
SEQ ID NO 177  
LENGTH: 1416  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (205) ..(1305)  
US-09-774-528-177

Alignment Scores:

Pred. No.: 2,51e-177 Length: 1416  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2\_copy\_148\_414 (1-267) x US-09-774-528-177 (1-1416)

```
QY 1 MetGUAIAHIAIAleuThrlleargLeuThraAspIleValIGluPherProheleuVal 20
DB 502 ATGAGGCTCAGACATTAATAAGTTAGTACCAATAAAGTAATTCCTTTTATGATT 561
QY 21 LeuLeuIleSerGlyGlyYHISCySeuLeuAlaleuValIGnglyValSerAspPheleu 40
DB 562 CTTTGGATTTTCGAGAGTCACTGTCTGTGGCATTTAGTTCAAGAGCTTTCAGATTTTCTG 621
QY 41 LeuLeuGlylySerSerleuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
DB 622 CTTCTTGGAAAGCTTTTGGACATAGACACAGGTGACATCTTGAACAAGGTGGCAAGAGA 681
QY 61 LeuSerleuIlelyHisProGluCySerThrMetSerGlyGlyYHISAlaIleGluHis 80
DB 682 CTTTCTTAATAATAACATCCAGAGTGTCCACCATGAGTGTGGAAAGCCATAGAACAT 741
QY 81 LeuAlaLySGnglyYAsnArgPheHisPheAspIleValProProleuHisIleAlaLyS 100
DB 742 TTGGCCCAACAAGAAATAGATTTCATTTGACATCAACCTCCCTTGACATCATGTATAA 801
QY 101 AsnCyAspPheSerPheThrGlyLeuGngIHisValThrAspIleIleIleMetLySlys 120
DB 802 AATTGTGATTTTCTTTTACTGGACTTCAACAGTTACTGATTAATAATATAGAAAG 861
QY 121 GluLySGluGluGlyIleGluLySGlyGngIleLeuSerSerAlaIleAspIleAla 140
DB 862 GAAAAAGAGGAAGTATGAGAAAGGGGCAAAATCTGTCTTCAGACAGCAACATTCGTGCC 921
QY 141 ThrValIGngIHisThMetAlaCySHISleuValLySArgThrHisArgAlaIleleuPhe 160
DB 922 ACGAGTACGACACAAATGATGATTCATTTGTAAGAAAGAACATCGGGCTATTCGTGTT 981
QY 161 CysLySGInArgAspLeuLeuProGInAsnAsnAlaValIleuValAlaSerGlyGlyVal 180
DB 982 TGTAGACAGAGAGACTGTGTACTCAAAATATGACATGCTGGTTCATCTGGTGTGTC 1041
QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGngIleLeuThrAsnAlaThrGInCySthr 200
DB 1042 GCAAGTAACTTCTATATCCGACAGAGCTCGAAATTTTAACAAACGCAACACAGTGCAC 1101
QY 201 LeuLeuCyProProProProArgLeuCyThrAspAsnGlyIleMetIleAlaIleAspGly 220
DB 1102 TTGTTGTGTCTCTCCCTCCAGACTATGACATGATTAATGACATTATGATTCAGAAATGAT 1161
QY 221 IleGluArgLeuArgAlaGlyLeuGngIleLeuHisAspIleGnglyIleArgArgGlu 240
DB 1162 ATTGAAAGACTACGCTGCTGGCTGGCATTTTACATGACATGAAAGGCAATCCGCTATGAA 1221
QY 241 ProLySGyProLeuGlyValAspIleSerLySGlyValIGlyGluAlaSerIlelyVal 260
DB 1222 CCAAAATGTCTCTTGGAGTACATATCAAAAGAAAGTTGAGAAAGCTTCCTCAAAAAGTA 1281
QY 261 ProGInLeuLyMetGluIle 267
DB 1282 CCACAAATTAATAATGAGATA 1302
```

RESULT 2  
US-10-067-443-23

; Sequence 23, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; TITLE OF INVENTION: SPINAL CORD, MP-1

FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 23  
LENGTH: 1526  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-067-443-23

Alignment Scores:  
Pred. No.: 2,85e-177 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2\_copy\_148\_414 (1-267) x US-10-067-443-23 (1-1526)

```
QY 1 MetGUAIAHIAIAleuThrlleargLeuThraAspIleValIGluPherProheleuVal 20
DB 1 ATGAGGCTCAGACATTAATAAGTTAGTACCAATAAAGTAATTCCTTTTATGATT 60
QY 21 LeuLeuIleSerGlyGlyYHISCySeuLeuAlaleuValIGnglyValSerAspPheleu 40
DB 61 CTTTGGATTTTCGAGAGTCACTGTCTGTGGCATTTAGTTCAAGAGCTTTCAGATTTTCTG 120
QY 41 LeuLeuGlylySerSerleuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
DB 121 CTTCTTGGAAAGCTTTTGGACATAGACACAGGTGACATCTTGAACAAGGTGGCAAGAGA 180
QY 141 ThrValIGngIHisThMetAlaCySHISleuValLySArgThrHisArgAlaIleleuPhe 160
DB 181 CTTTCTTAATAATAACATCCAGAGTGTCCACCATGAGTGTGGAAAGCCATAGAACAT 240
QY 161 LeuSerleuIlelyHisProGluCySerThrMetSerGlyGlyYHISAlaIleGluHis 80
DB 181 CTTTCTTAATAATAACATCCAGAGTGTCCACCATGAGTGTGGAAAGCCATAGAACAT 240
QY 81 LeuAlaLySGnglyYAsnArgPheHisPheAspIleValProProleuHisIleAlaLyS 100
DB 241 TTGGCCCAACAAGAAATAGATTTCATTTGACATCAACCTCCCTTGACATCATGTATAA 300
QY 121 GluLySGluGluGlyIleGluLySGlyGngIleLeuSerSerAlaIleAspIleAla 140
DB 301 AATTGTGATTTTCTTTTACTGGACTTCAACAGTTACTGATTAATAATAGAAAG 360
QY 101 AsnCyAspPheSerPheThrGlyLeuGngIHisValThrAspIleIleIleMetLySlys 120
DB 301 AATTGTGATTTTCTTTTACTGGACTTCAACAGTTACTGATTAATAATAGAAAG 360
QY 121 GluLySGluGluGlyIleGluLySGlyGngIleLeuSerSerAlaIleAspIleAla 140
DB 121 GluLySGluGluGlyIleGluLySGlyGngIleLeuSerSerAlaIleAspIleAla 140
QY 141 ThrValIGngIHisThMetAlaCySHISleuValLySArgThrHisArgAlaIleleuPhe 160
DB 361 GAAAAAGAGGAAGTATGAGAAAGGGGCAAAATCTGTCTTCAGACAGCAACATTCGTGCC 420
QY 161 CysLySGInArgAspLeuLeuProGInAsnAsnAlaValIleuValAlaSerGlyGlyVal 180
DB 421 ACGAGTACGACACAAATGATGATTCATTTGTAAGAAAGAACATCGGGCTATTCGTGTT 480
QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGngIleLeuThrAsnAlaThrGInCySthr 200
DB 481 TGTAGACAGAGAGACTGTGTACTCAAAATATGACATGCTGGTTCATCTGGTGTGTC 540
QY 201 LeuLeuCyProProProProArgLeuCyThrAspAsnGlyIleMetIleAlaIleAspGly 220
DB 601 TTGTTGTGTCTCTCCCTCCAGACTATGACATGATTAATGACATTATGATTCAGAAATGAT 660
QY 221 IleGluArgLeuArgAlaGlyLeuGngIleLeuHisAspIleGnglyIleArgArgGlu 240
DB 661 ATTGAAAGACTACGCTGCTGGCTGGCATTTTACATGACATGAAAGGCAATCCGCTATGAA 720
```



```
QY 241 ProlyscProLeuGlyValAspIleSerIleGlyValAlaSerIleIleVal 260
DB 721 CCAAAATGCTCTTGGAGTAGACATATCAAAAAGAGTTGAGAGAGCTTCCATTAAGA 780
QY 261 ProGlnLeuIleMetGluIle 267
DB 781 CCACAAATTAATAATGAGATA 801

RESULT 3
US-10-067-443-1
; Sequence 1, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-067-443-1

Alignment Scores:
Pred. No.: 5,23e-177 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-1 (1-2197)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGlnPheProPheLeuVal 20
DB 672 ATGAGAGGCTCAGTCACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 731
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 732 CTTTGGATTTCTGAGAGTCACTGCTGTTGGCAATTGCAAGAGTTTCAGATTTTCTG 791
QY 41 LeuLeuGlyIleSerIleuAspIleAlaProGlyAspMetLeuAspIleValAlaArg 60
DB 792 CTTCTTGGAAAGCTTTTGGACATACCAAGTGCATCTTGGACAGAGTGCAGAGAGA 851
QY 61 LeuSerLeuIleIleValHisProGlyCysSerThrMetSerGlyIleValAlaIleGluHis 80
DB 852 CTTTCTTAATTAATAACATCCAGAGTGTCCACCACTGAGTGGGAGAAAGCCATAGAACAT 911
QY 81 LeuAlaIleGlnGlyAsnArgPheHisPheAspIleIleValProProLeuHisHisAlaIle 100
DB 912 TTGGCCAAACAGAGAAATGATTTGATTTGACATCAACCTCCCTTGATCATCTAA 971
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleIleMetIleVal 120
DB 972 AATTGATTTTCTTTTACTGAGATTTCAACGTTACTGATTAATAATTAATGAAGAAG 1031
QY 121 GlnIleGlnGlnGlyIleGlnIleGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 1032 GAAAAAGAGAGGATATTGAGAGAGGAGCAAAATCTGTCTTGACAGACATTTCTGCC 1091
QY 141 ThrValGlnHisThrMetAlaCysHisIleuValIleValGlnHisArgAlaIleLeuPhe 160
```

```
DB 1092 ACAGTACAGACACAAATGAGCATGTCTTGTGAAAAGAACACATCGGGCTATTCTGTTT 1151
QY 161 CylGlyGlnArgAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyVal 180
DB 1152 TGTAAAGACAGAGACTGTTTAACTCTCAAAATTAAGCAGTCTGTTTCATCTGTGTGTTC 1211
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200
DB 1212 GCAGTACTTCTATATCCGACAGCTCTGGAAATTTTAAACAAACGACACAGTGCATCT 1271
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIlePaangly 220
DB 1272 TTGTTGTGCTCTCTCCAGACATGACATGATTAATGCACTTATGATTCATGGAATGGT 1331
QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgIle 240
DB 1332 ATTTGAAAGACTAGTCTGCTGGCTTGGCAATTTAACAAGACATGAGAGCATCCCTATGAA 1391
QY 241 ProlyscProLeuGlyValAspIleSerIleGlyValAlaSerIleIleVal 260
DB 1392 CCAAAATGCTCTTGGAGTAGACATATCAAAAAGAGTTGAGAGAGCTTCCATTAAGA 1451
QY 261 ProGlnLeuIleMetGluIle 267
DB 1452 CCACAAATTAATAATGAGATA 1472

RESULT 4
US-10-067-443-21
; Sequence 21, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED
; TITLE OF INVENTION: SPINAL CORD, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-21

Alignment Scores:
Pred. No.: 9.58e-174 Length: 1387
Score: 1358.50 Matches: 266
Percent Similarity: 91.10% Conservative: 0
Best Local Similarity: 91.10% Mismatches: 1
Query Match: 98.09% Indels: 25
DB: 4 Gaps: 1

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-21 (1-1387)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnIleValGlnPheProPheLeuVal 20
DB 465 ATGAGAGGCTCAGTCACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 524
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 525 CTTTGGATTTCTGAGAGTCACTGCTGTTGGCAATTGTTCAAGAGTTTTCAGATTTTCTG 584
QY 41 LeuLeuGlyIleSerIleuAspIleAlaProGlyAspMetLeuAspIleValAlaArg 60
DB 585 CTTCTTGGAAAGCTTTTGGACATACCAAGTGCATCTTGGACAGAGTGCAGAGAGA 644
QY 61 LeuSerLeuIleIleValHisProGlyCysSerThrMetSerGlyIleValAlaIleGluHis 80
```

|||||  
Db 645 CTTTCCTTAATAAATCAATGAGTGTCCACCATGAGTGGGAAAGCCATAGAGCAT 704  
Qy 81 LeuAlaIyGInGlyAserAryPheHisPheAserPileYsProProLeuHisHisAlaIyS 100  
Db 705 TTGGCCAAACAGGAATAATGATTTTCATTTTGACATCAACCTCCCTTGATCATGCTAAA 764  
Qy 101 AserCysAserPheSerPheThGlyLeuGInHisValIThrAplysIleIleMetIySlys 120  
Db 765 AATTTGATTTTTCTTTTACTGAGCTTCAACAGCTTACTGATTAATAATTAATGAAGAAAG 824  
Qy 121 GluIyGInGlyIle-----GluIyGInGlyIleIleuSerSeralA 126  
Db 825 GAAAAGAGGAAGATATTTCTAATAGTAAAGTTGAACAGATAAATATCTCGATTC 884  
Qy 127 -----GluIyGInGlyIleIleuSerSeralA 135  
Db 885 TGCCATAAATAATAGCTCTCATTTCTGACAGTATGAGAAAGGGCCAAATCCTGCTTCAGCA 944  
Qy 136 AlaAserPileAlaIleThrValGInHisThrMetAlaCysHisIleuValIySArgThrHis 155  
Db 945 GAGACATGCTGCGACAGTACAGACACAATGCGATGTCATCTGTGAAAGAACACACT 1004  
Qy 156 ArgAlaIleLeuPheCysIyGInAryAserAplyLeuProGInAserAserAlaValIleuVal 175  
Db 1005 CGGCGTATTCGTTTGTATAGCAGAGAGACTGTTTACCTCAAAATATATGACATGCTGTT 1064  
Qy 176 AlaSergIyGlyValAlaSerAserPheIyIleAryAryAlaLeuGInIleuThrAen 195  
Db 1065 GCATCTGGTGGTGGTGGCAATTAATCTTATATCCGACAGCTCTGGAAATTTTAAACAAAC 1124  
Qy 196 AlaThrGInCysThrLeuLeuCysProProAryGleuCysThrAserAserIyIleMet 215  
Db 1125 GCAACACAGTGCATCTTGTGTGTCTCTCCCAACATATGACATATATGACATTAAG 1184  
Qy 216 IleAlaIlePAsnGlyIleGInAryGleuAryAlaGlyIleuGlyIleuHisAserPileGlu 235  
Db 1185 ATTCGATGGAATGATGATGAAAGACTACGTGCGCTTGCGCATTTTACATGACATAGAA 1244  
Qy 236 GlyIleAryIyGInProIyCysProLeuGlyValAserPileSergIyValIyGlyIu 255  
Db 1245 GGCATCCGCTATGAACCAAAATGTCTCTTGAGATAGACATATCAAAAGAGTGGAGAA 1304  
Qy 256 AlaSergIleYsValProGInleuIySmetGluIle 267  
Db 1305 GCTTCCATTAAGATACCAATTAATAATGAGATA 1340  
RESULT 5  
US-10-067-443-20  
; Sequence 20, Application US/10067443  
; Patent No. 6642041  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; TITLE OF INVENTION: SPINAL CORD, ME-1  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 20  
; LENGTH: 14364  
; TYPE: DNA  
; ORGANISM: homo sapiens  
; US-10-067-443-20  
Alignment Scores:  
Pred. No.: 1,34e-99 Length: 14364  
Score: 824.50 Matches: 240

Percent Similarity: 26.73% Conservative: 0  
Best Local Similarity: 26.73% Mismatches: 2  
Query Match: 59.53% Indels: 657  
DB: 4 Gaps: 4  
US-10-649-273-2\_copy\_148\_414 (1-267) x US-10-067-443-20 (1-14364)  
Qy 1 MetGluAlaHisAlaLeuThrIleAryGleuThrAserIySValIyGInPheProPheLeuVal 20  
Db 10845 ATGAGGCTCATAGCACTTACTATAGGTTGACCAATTAAGTGAATTTCTTTTATGTT 10904  
Qy 21 LeuLeuIleSergIyGlyHisCysLeuLeuAlaLeuValGInGlyValIleAserPheLeu 40  
Db 10905 CTTTGAATTTCTGAGAGTCACTGTGTTGGCATTAGTTCAAGAGTTTCAAGATTTTCG 10964  
Qy 41 LeuLeuGlyIySergIyAryHisCysLeuLeuAlaLeuValIySmetIySValIySVal 57  
Db 10965 CTTTCATGGAAGCTTTTGACATAGCACAGGATGACATCTTGACAAAGGT-AATTAAGA 11023  
Qy 57 ----- 57  
Db 11024 TTAATTTCTCATTTCTTTTGTATGTTGTCATTTCATTAAGTAACTGAATATGATGTC 11083  
Qy 57 ----- 57  
Db 11084 TACCACTTCACTTAATATTTTGAATTTTATCTAGTAAACTGAAAAATTCACAT 11143  
Qy 57 ----- 57  
Db 11144 ANGTGAGAAAAAATAGAAAGATAGTACACAAATTTATATCTTACGCTTTCTTAATA 11203  
Qy 57 ----- 57  
Db 11204 AAATGTAAGAGTTCATATCTGTACATTAAGGCTGAATAGTTGACATACATTAAG 11263  
Qy 57 ----- 57  
Db 11264 TATTTGGCCAAATATGATATGTAAGAAAGCTGCTTGTAACTAATACATGCAAAAAA 11323  
Qy 57 ----- 57  
Db 11324 GGTAAATTAAGAAATATATATATATATTAATTAACATTAAGACATTAAGATGCAATGACAGAA 11383  
Qy 57 ----- 57  
Db 11384 TTAATCACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11443  
Qy 57 ----- 57  
Db 11444 AATATCTACAGAGCTTACCTATATATAGAAAAACAAACAAACAAACAAACAAACAACTG 11503  
Qy 57 ----- 57  
Db 11504 CTTCCACAGTGAATTAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTA 11563  
Qy 57 ----- 57  
Db 11564 AAGCAGTATTTGCACTTCAAGCCCATTTTCCAAACCAATTAAGAAAGCAAAACATAGACAG 11623  
Qy 57 ----- 57  
Db 11624 GGGCAGTATTTGGCTTTATTTGTTGGGTCAATTAAGAAACAGGTTGTCTGCTTAC 11683  
Qy 57 ----- 57  
Db 11684 TGAATATCAGCTATAGTATATATTTGCCAAAGATAGCATTTTATTCATTCAGGGGTT 11743  
Qy 57 ----- 57  
Db 11744 TTTTGTGTTAGTAAATTTCAATTTATTTCTTTGACATCTTTGCTTTCACAGATTTT 11803  
Qy 58 -----AlaAryAryIleuSergIle 64  
|||||

```

Db 11804 AATTTTAACTCTAAATAATATGTTCTTGTAGTGGGCAAGACTTCTTAATA 11863
QY 65 LysHisProGluCysSerThreSerGlyValAlaIleGluHisIleuAlaIleuGln 84
Db 11864 AAACATCCAGAGTGTCCACCAATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACA 11923
QY 85 GlyAsnArgPheHisPheAspIleLeuProPheHisIleAlaIleuAsnCysAspPhe 104
Db 11924 GGAATATATTTCACTTTGACATCAACACTCCCTTGACATCACTTAAATATGTGATTTT 11983
QY 105 SerPheThrGlyLeuGlnHisValIleuThrAspIleIleMetLeuGlyValGluGlu 124
Db 11984 TCTTTTACTGACTTCAACACCTTACTGATTAATATATGATAAAGGAAAGGAA 12043
QY 124 ----- 124
Db 12044 GGTATATTTCTAATATGTAAGTGAACAGATAAATATCTGTGATGTGCTAAATAA 12103
QY 125 -----GlyIleGlyValGlnIleuSerSerAlaIleAspIleAl 139
Db 12104 GCTGCTATTTCTGCAGGTATTTGAGAAAGGCAAACTCTGTCTTCAAGCAGACATTC 12163
QY 139 AlaIleThrValGlnHisIleuMetAlaCysHisIleuValIleuArgThrHisArgAlaIle 159
Db 12164 TGCCACAGTACAGCACACAAATGGCATGTCATCTTGAAAGAAAGACACATCGGGCTATTCT 12223
QY 159 uPheCysIleuGlnHisIleuAspLeuLeuProGlnHisIleuAlaValIleu----- 174
Db 12224 GTTTTGTAGGACAGAGACTGTGTACTCTCAAAATATATGACATGCTGTAAGTTTATCTC 12283
QY 174 ----- 174
Db 12284 ATTTTATATGTAATATGTAACACTTGGCAATATGTAATCTTTTCCACAGACTTGA 12343
QY 174 ----- 174
Db 12344 TGTTAGAGTGAACAGATCTTATGCTTATGCTAGCCCTGACAGATGAATATATG 12403
QY 174 ----- 174
Db 12404 GATGAAAGACTAACAGCCATTTCTGTACTAGTTGGTAGCTTATGAGCAGCTGTGA 12463
QY 174 ----- 174
Db 12464 TAGCTTATAGGACATATAGTCTAATTTGCAATCTTCTGTGATTTAAAGAGGCTT 12523
QY 174 ----- 174
Db 12524 ACAATAAAGAAAGTAATGACATACGTCTATCACTATTTTGAAGAAATAGTGTGATTT 12583
QY 174 ----- 174
Db 12584 CCTTCATCTTGAATGAATCCCTTGTGTTGTTGTTTTTAATAGCCAGTCAAAATTT 12643
QY 174 ----- 174
Db 12644 AGCAGTGGAGGTGTATTCCAACTTTCGTGACACTAAATGTTGATTAAGTTCTGATATC 12703
QY 174 ----- 174
Db 12704 CACTATATTTGACAGCCAAATCCCTTAATTTGCTTAAAGCCTTGACAAACATCTCT 12763
QY 174 ----- 174
Db 12764 GTTTAAGTATCTTAATCTTATTCATTTAAATAATTAATAAGTGAAGAAATGTT 12823
QY 174 ----- 174
Db 12824 TAAATGTAGTAATTCATATGATGAATTTTACATGATATCAAGAAATATTTTTCAGA 12883
QY 174 ----- 174
Db 12884 GTTATGTAGTAATAATGACAAATAATTAATAATTTCAAGGCTTAAATAATAGTCTATG 12943

```

```

QY 174 ----- 174
Db 12944 TTGAATATATTAATAATAATATTTAGATGAAGGTTGGAAGAAATATACAAAATGCT 13003
QY 174 ----- 174
Db 13004 AGTAATGTTGATAGTATTAAGAAATTAATTAATTTTCTTCCAAATTTTATATAC 13063
QY 174 ----- 174
Db 13064 ATAGATATGTCATCTGCCATTACCATCTCAAAATGGATAGTTATATGTTTAATG 13123
QY 174 ----- 174
Db 13124 CTGATATTTTCTCCAGTTTAATTAAGACGCTTGTCATATCCATATATGATGTTATT 13183
QY 175 -----ValAlaSerGlyValAlaSerAspPheThrIleu 187
Db 13184 TTGTTTCTCAATTCCTTCAGGTTGCACTGTGTGTGTCGCAAGTATCTTATATCCG 13243
QY 187 gAlaIleuGlnIleuThrAsnAlaThrGlnCysThrLeuCysProProAla 207
Db 13244 CAGAGCTGTGAATTTTAACAAACGCAACAGTGCACCTTGTGTCTCTCCAG 13303
QY 207 gIleuCysThrAspAsnGlyIleuMetIleuAlaTrp----- 218
Db 13304 ACTATGCACTGATTAATGGAATTAATGATTCATGTGAAGCAGATATACGTGCTTAC 13363
QY 218 ----- 218
Db 13364 TCATATCTATGTAATATTAATTCGCATTTTATCATACTAAGCTTCTTCAGATCT 13423
QY 219 -----AsnGlyIleGluArgLeuArg 226
Db 13424 TGAGCTATGATTTTATTTAATGCTTCTTATTTAGATGATGATTTGAAGCTACGTG 13483
QY 226 lAglyLeuGlnIleuHisAspIleGluGlyIleuArgTyrGluProLys 242
Db 13484 CTGGCTTGCGCATTTTATCATGACATGAAGGCAATCCGCTATGAACCAAG 13533

RESULT 6
US-09-596-002-38
; Sequence 38, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, B.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAYELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 38
; LENGTH: 94750
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: incyte template ID No. 6632636 38
; PUBLICATION INFORMATION:
US-09-596-002-38

Alignment Scores:
Pred. No.: 3,216-26 Length: 94750
Score: 308.50 Matches: 87
Percent Similarity: 49.61% Conservative: 40
Best Local Similarity: 33.98% Mismatches: 101
Query Match: 22.27% Indels: 29

```



```
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 884
/ LENGTH: 1059
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884

Alignment Scores:
Pred. No.: 3,67e-29      Length: 1059
Score: 291.50           Matches: 87
Percent Similarity: 50.41%  Conservative: 37
Best Local Similarity: 35.37%  Mismatches: 99
Query Match: 21.05%      Indels: 23
DB: 4                  Gaps: 9

US-10-649-273-2_COPY_148_414 (1-267) x US-09-252-991A-884 (1-1059)

QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPheLeu 19
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      711 ATGGAAGGGCACCTGCTGGCCGCGATGCTGGAAGAGACACACCGCGGTTCCCGTTCCGTC 652

QY      20 ValLeuLeuIleSerGlyYhiScyLeuLeuAlaLeuValGlnGlyValSerAspHe 39
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      651 GCCCTGCTGTTCCGGCGGTACACCCAGTTGTGCGGCGGAGACGGTATCGGCCGCGTAC 592

QY      40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      591 CAGTTGCTTGGCGAATCGGTGACATGCCCGCGGAGACCTTGCACAAAGCCGCCAAG 532

QY      60 ArgLeuSerLeuIleLysAspProGluCysSerThrMetSerGlyLysAlaIleGlu 79
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      531 CTGATCGGCGCTG---GGCTATCCC-----GGTGGTCCGGAATCCGC 493

QY      80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAla 99
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      492 CGCTGCGGAGCGCGGACACTCTGCGCTTCGTTCCCGCGCGCATGACCCGATCGC 433

QY      100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      432 CCCGCGCTGACCTTCAAGCTTCAAGCGGCTCAAGACCTTACCTGAC--ACTGGCAG 376

QY      120 LysGluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaSerIleAla 139
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      375 CGTTGCGTCAAGCGCGGCGAGCAGACGAGCAG-----ACCGCTGCAGACTCCGC 325

QY      140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      324 CTGGCGTTCCAGACCGCGGTGTGAGACCTGCTGATCAAGTGCCTGCGGCTG--- 268

QY      160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyLys 179
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      267 -----AAGCAGACCGGCTG-----AAGAAC-----CTGGTATCGCGCGCGGT 229

QY      180 ValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      228 GTCAAGCCCAACAGCGGCTCGCAGCGGCTGGAAGAGATCTCGGGAATGAAGGGG 169

QY      200 ThrLeuLeuCysProProAlaArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsn 219
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      168 CAGGTGTTTCAACGCCCGCGCTTGCACCGACATGCGGAGTATGATCGCTACGCC 109

QY      220 GlyIleGluArgLeuAlaGlyIleGluGlyIleLeuHisAspIleGluGlyIleArgTyr 239
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      108 GGTGCTCAGGCGCTGCTCGCGGC-----CAGCATGACGCGCCGCGATGACCGCTC 58

QY      240 GluProLysCysProLeu 245
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      57 CAGCGCGCTGCGCGCATG 40
```

```
RESULT 9
US-09-252-991A-801
/ Sequence 801, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfeld et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-07-27
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 801
/ LENGTH: 1206
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-801

Alignment Scores:
Pred. No.: 4,55e-29      Length: 1206
Score: 291.50           Matches: 87
Percent Similarity: 50.41%  Conservative: 37
Best Local Similarity: 35.37%  Mismatches: 99
Query Match: 21.05%      Indels: 23
DB: 4                  Gaps: 9

US-10-649-273-2_COPY_148_414 (1-267) x US-09-252-991A-801 (1-1206)

QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLys---ValGluPheProPheLeu 19
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      514 ATGGAAGGGCACCTGCTGGCCGCGATGCTGGAAGAGACACACCGCGGTTCCCGTTCCGTC 573

QY      20 ValLeuLeuIleSerGlyYhiScyLeuLeuAlaLeuValGlnGlyValSerAspHe 39
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      574 GCCCTGCTGTTCCGGCGGTACACCCAGTTGTGCGGCGGAGACGGTATCGGCCGCGTAC 633

QY      40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      634 CAGTTGCTTGGCGAATCGGTGACATGCCCGCGGAGACCTTGCACAAAGCCGCCAAG 693

QY      60 ArgLeuSerLeuIleLysAspProGluCysSerThrMetSerGlyLysAlaIleGlu 79
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      694 CTGATCGGCGCTG---GGCTATCCC-----GGTGGTCCGGAATCCGC 732

QY      80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAla 99
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      733 CGCTGCGGAGCGCGGACACTCTGCGCGCTTCGTTCCCGCGCGCATGACCCGATCGC 792

QY      100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      793 CCCGCGCTGACCTTCAAGCTTCAAGCGGCTCAAGACCTTACCTGAC--ACTGGCAG 849

QY      120 LysGluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaSerIleAla 139
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      850 CGTTGCGTCAAGCGCGGCGGACAGCAGCAG-----ACCGCTGCAGACTCCGC 900

QY      140 AlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeu 159
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      901 CTGGCGTTCCAGACCGCGGTGTGAGACCTGCTGATCAAGTGCCTGCGGCTG--- 957

QY      160 PheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyLys 179
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      958 -----AAGCAGACCGGCTG-----AAGAAC-----CTGGTATCGCGCGCGGT 996

QY      180 ValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199
       |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      997 GTCAAGCCCAACAGCGGCTCGCAGCGGCTGGAAGAGATCTCGGGAATGAAGGGG 1056
```

[illegible]

Db	552821	GCTGTGTTGATGCGGGTGGCCACACTCAATTAGTGCCTGTGCAAGTGTAGGAAATAT	552762
Qy	40	LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg	59
Db	552761	GAACTGATGAGAGAAATCTATTGATATATCTGCTGGCAAGCCTTTGATTAACACGAAAA	552702
Qy	60	ArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyValLysAlaIleGlu	79
Db	552701	TTACTTGGACTA---GATTATCCA-----CGTGGCCGGCACCTTCT	552663
Qy	80	HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAla	99
Db	552662	CGTTTAGCGGAAAAAGTAGACCGCAATCGTTTCACATTTCCACGTCCAAATGACAGATCGT	552603
Qy	100	LysAsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys	119
Db	552602	GCAGCGCTTGATTTTATGTTTTTCTGTGTTTAAACAAATTTGCCGCAATACAGTTAATCA	552543
Qy	120	LysGlnLysGlnGluGlnGly-----IleGlnLysGlnIleLeuSerSerAlaAlaAsp	137
Db	552542	GCAATTTAAAAACGAGGCGCAACTGTATAGCAA-----ACTTAAAGCAGAT	552488
Qy	138	IleAlaIaIaIaIaIaIaGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla	157
Db	552497	ATTGCTTATGCTTTCCAAAGATCGGTGGTGAATCTTGCC-----	552456
Qy	158	IleLeuPheCysLysGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSer	177
Db	552455	--ATTAAATGTAAAG--CGTGCAATTTGAAAGAAACAGCGTTATTAACGTTTAGATTGCG	552402
Qy	178	GlyGlyValAlaIaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThr	197
Db	552401	GGAGGGGAGAGGCAATTAATAAACTCCGAGAAACGCTTGCGCACTTAATGCAAAATTTA	552342
Qy	198	GlnCysThrLeuLeuCysAspProProArgLeuCysThrAspAsnGlyIleMetIleAla	217
Db	552341	GGTGGCGAAGGTGTTTATCTCAACCTCAATTTGTGTACAGATATATGTCGATGATTGCT	552282
Qy	218	TrpAsnGlyIleGluArgLeuArgAlaGly	227
Db	552281	TACACAGGTTTTTTCGTTTAAACAAAGGT_552252	
RESULT 11			
US-09-643-990A-1/c			
; Sequence 1, Application US/09643990A			
; Patent No. 6528289			
GENERAL INFORMATION:			
APPLICANT: Robert D. Fleischmann			
Mark D. Adams			
Owen White			
Hamilton O. Smith			
J. Craig Venter			
TITLE OF INVENTION: The Nucleotide sequence of			
the Haemophilus influenzae Rd Genome, Fragments			
Thereof, and Uses Thereof			
NUMBER OF SEQUENCES: 1			
CORRESPONDENCE ADDRESS:			
ADDRESSER: Human Genome Sciences, Inc.			
STREET: 9410 Key West Avenue			
CITY: Rockville,			
STATE: MD			
COUNTRY: USA			
ZIP: 20850			
COMPUTER READABLE FORM:			
MEDIUM TYPE: 3 1/2 inch diskette			
COMPUTER: Dell Pentium			
OPERATING SYSTEM: MS DOS v6.22			
SOFTWARE: ASCII Text			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/09/643,990A			
FILING DATE: 23-Aug-2000			
CLASSIFICATION: <Unknown>			
PRIOR APPLICATION DATA:			
/			

APPLICATION NUMBER: 08/487,429  
 FILING DATE: 1995-06-07  
 APPLICATION NUMBER: 08/426,787  
 FILING DATE: 1995-04-21  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kenley K. Hoover  
 REGISTRATION NUMBER: 40,302  
 REFERENCE/DOCKET NUMBER: PB186P1C1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 301-610-5790  
 TELEFAX: 310-309-8439  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1830121 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-643-990A-1

Alignment Scores:  
 Pred. No.: 1.68e-23 Length: 1830121  
 Score: 289.50 Matches: 82  
 Percent Similarity: 48.70% Conservative: 30  
 Best Local Similarity: 35.65% Mismatches: 95  
 Query Match: 20.90% Indels: 23  
 Gaps: 7  
 DB: 4

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-643-990A-1 (1-1830121)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeu---ThrAsnLysValGluPheProPheLeu 19  
 DB 552881 ATGGAAGGCGATTACTTGGCCCAATGCTTATGACAACTTACCGGCACTTCTTTCTT 552822  
 QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39  
 DB 552821 GCCTGTTGGTATCGGGTGGCCACACTCAATTAGTCGTGCATGGTGTGGAATAATAT 552762  
 QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59  
 DB 552761 GAAGTGTAGAGAAATCTATTGAATGATCTCTCGCGGAAGCTTGTGAATAAACAAGCAA 552702  
 QY 60 ArgLeuSerLeuIleLysHisArgProGlyCysSerThrMetSerGlyGlyValAlaIleGlu 79  
 DB 552701 TTTACTTGACATA---GATTATCCA-----GGTGGCGCGCACTTCT 552663  
 QY 80 HisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAla 99  
 DB 552662 CGTTTACGGAAAAAGAGTACGCCCAATGCTTTCACATTTCCACGTCCCAATGACAGATGCT 552603  
 QY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLys 119  
 DB 552602 GCAGGCGCTGATTTAGTTTCTGCTTAAACAATTGCGCGCAAAATPACAGTTAATCAA 552543  
 QY 120 LysGlnLysGlnGluGly-----IleGlnLysGlnGlnIleLeuSerSerHisAlaAsp 137  
 DB 552542 GCAATTAAAAACGAGCGCAATGATAGAGCA-----ACTAAAGGCAT 552498  
 QY 138 IleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAla 157  
 DB 552497 ATTGCTTATGCTTCCAAAGATGCGGTGTGATACTCTTGCC----- 552456  
 QY 158 IleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaValLeuValAlaSer 177  
 DB 552455 ---ATTAAATGTAG---CGTGCATTGAAAGAAACAGGCTATAAAGCTTTAGTGTGCG 552402  
 QY 178 GlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThr 197  
 DB 552401 GAGGGGTGGCGCGCAATAAATACTCGAGAAACGCTTGGCGCACTTATGCAAAATTTA 552342  
 QY 198 GlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAla 217  
 DB 552341 GGTGGCGAAGGTTTATCTCAACCTCAATTGTGTACAGTAATGGTGCAGATGATGCT 552282

QY 218 TrpAsnGlyIleGluArgLeuArgAlaGly 227  
 DB 552281 TACACAGGTTTCTTACGTTTAAACAAGGT 552252

RESULT 12  
 US-09-902-540-6612  
 Sequence 6612, Application US/09902540  
 Patent No. 6833447  
 GENERAL INFORMATION:  
 APPLICANT: Goldman, Barry S.  
 APPLICANT: Hinkle, Gregory J.  
 APPLICANT: Slater, Steven C.  
 APPLICANT: Wiegand, Roger C.  
 TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 FILE REFERENCE: 38-10(15849)B  
 CURRENT APPLICATION NUMBER: US/09/902,540  
 PRIOR FILING DATE: 2001-07-10  
 PRIOR APPLICATION NUMBER: 60/217,883  
 NUMBER OF SEQ ID NOS: 16825  
 SEQ ID NO 6612  
 LENGTH: 996  
 TYPE: DNA  
 ORGANISM: Myxococcus xanthus  
 US-09-902-540-6612

Alignment Scores:  
 Pred. No.: 6.44e-28 Length: 996  
 Score: 282.00 Matches: 86  
 Percent Similarity: 48.93% Conservative: 28  
 Best Local Similarity: 36.91% Mismatches: 87  
 Query Match: 20.36% Indels: 32  
 Gaps: 9  
 DB: 4

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-09-902-540-6612 (1-996)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysVal---GluPheProPheLeu 19  
 DB 334 CTGAGGGCGCACCTGCTGCGCCATCGCTGTGAGAGTGGCCCGGAGCGGCTTCTT 393  
 QY 20 ValLeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspHe 39  
 DB 394 GGGCTGCTGCTTCCGGGGGACACACAGCTTACAGAGTGCAGGCTTACGGGCAGTAC 453  
 QY 40 LeuLeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArg 59  
 DB 454 CGGCTGTGGGAGACAGCGCGGACGCGGCGGAGGACATATGACAAAGACCGCTGCG 513  
 QY 60 ArgLeuSerLeuIleLysHisArgProGlyCysSerThrMetSerGlyGlyValAlaIleGlu 79  
 DB 514 ATCTCTGCGCTG---CCGATCCG-----GGTGGGAGCGCCATCGAC 552  
 QY 80 HisLeuAlaLysGlnGlyAsn-----ArgPheHisPheAspIleLysProPro 95  
 DB 553 CAGTTGGCGACAGAGGAAACCGGAGGCCATCCCTT-----CGGCGC 597  
 QY 96 LeuHisHisAlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLys 115  
 DB 598 GCGCTCGCGCGGACAACTTGCACCTGCTTCTCGGCTTAAAG-----ACG 645  
 QY 116 IleIleMetLysLysGlnLysGlnGlyIleGlnLysGlnIleLeuSerSerAla 135  
 DB 646 GCGGCTGCACACAGTGCAGAGGACCGGCGCGGAGGCGGAGCGCGCTG----- 696  
 QY 136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155  
 DB 697 GGGATTTGTGCGCTCTTCCAGAGGAGCGGTGGCGAGCTGCTGTGCAAGAG----- 750  
 QY 156 ArgAlaIleLeuPheCysLysGlnArgAspLeuProGlnAsnAlaValLeuVal 175  
 DB 751 -----CTGTGGCGCGCGCGCGCGCTTG-----GCCACAAAGCATTTGGTG 792







Thu Feb 17 17:37:57 2005

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 18:03:13 ; Search time 534 Seconds  
(without alignments)  
2950.230 Million cell updates/secTitle: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHALTRLRNKKVPPFLV.....DISKVGESAIKVPQMKMEI 267Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0Searched: 5378673 seqs, 2950229984 residues  
Total number of hits satisfying chosen parameters: 10757346Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summariesCommand line parameters:  
-MODE=frame+ p2n.model -DEV=x1h  
-Q=/cgn2\_1/USPTO.epool/US10649273/rnatc.14022005.114706.16497/app.query.fasta\_1.1429  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=mpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdt -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10649273@cgn\_1.1.1053@rnatc.14022005.114706.16497  
-NCPU=6 -ICPU=3 -NO MMAP -LARGECQUERY -NEG SCORES=0 -WAIT -DSPLOCH=100  
-LONGLOG -DEV TIMEOUT=110 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7Database : Published Applications\_NA: \*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US10F\_NEW\_PUB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	1416	17	US-10-120-988-177
2	1385	100.0	1526	14	US-10-067-443-23
3	1385	100.0	1526	17	US-10-649-273-23
4	1385	100.0	1526	17	US-10-651-722-23
5	1385	100.0	2197	14	US-10-067-443-1
6	1385	100.0	2197	17	US-10-649-273-1
7	1385	100.0	2197	17	US-10-651-722-1
8	1358.5	98.1	1387	14	US-10-067-443-21
9	1358.5	98.1	1387	17	US-10-649-273-21
10	1358.5	98.1	1387	17	US-10-651-722-21
11	1348	97.3	1245	14	US-10-012-140-6
12	1348	97.3	1820	14	US-10-012-140-4
13	1204	86.9	2208	17	US-10-094-749-400
14	1204	86.9	2890	18	US-10-723-860-7447
15	824.5	59.5	14364	14	US-10-067-443-20
16	824.5	59.5	14364	17	US-10-649-273-20
17	824.5	59.5	14364	17	US-10-651-722-20
18	468	33.8	371	18	US-10-430-201-3118
19	468	33.8	371	18	US-10-430-201-3119
20	422.5	30.5	1917	17	US-10-424-539-66417
21	359	25.9	1628	18	US-10-437-963-11249
22	342	24.7	756	14	US-10-081-051-8
23	342	24.7	4360	14	US-10-081-051-2
24	321.5	23.2	1146	17	US-10-282-122A-14674
25	314.5	22.7	936	17	US-10-282-122A-8315
26	308.5	22.3	1044	17	US-10-282-122A-26972
27	308.5	22.3	94750	17	US-10-672-787-38
28	304.5	22.0	1032	17	US-10-282-122A-31043
29	301.5	21.8	1000	18	US-10-343-561-50
30	291.5	21.0	1026	9	US-09-815-242-7701
31	291.5	21.0	1026	17	US-10-282-122A-30016
32	289.5	20.9	1029	9	US-09-815-242-6546
33	289.5	20.9	1029	17	US-10-282-122A-22020
34	289.5	20.9	1830121	17	US-10-329-670-1
35	289.5	20.9	1830121	18	US-10-158-865-1
36	285	20.6	372	18	US-10-357-930-61088
37	280.5	20.3	1020	17	US-10-282-122A-32254
38	279.5	20.2	1014	9	US-09-815-242-9682
39	279.5	20.2	1014	17	US-10-282-122A-39301
40	279.5	20.2	1014	17	US-10-282-122A-41977
41	277.5	20.0	1007	17	US-10-282-122A-19220
42	277	20.0	927	17	US-10-282-122A-25354
43	275.5	19.9	1023	17	US-10-282-122A-31809
44	275.5	19.9	1014	9	US-09-815-242-6207
45	275.5	19.9	1014	17	US-10-282-122A-20511

## ALIGNMENTS

RESULT 1  
US-10-120-988-177  
; Sequence 177, Application US/10120988  
; Publication No. US20030219745A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Ren, Feiyun  
; APPLICANT: Wang, Dunhui  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 802CON  
; CURRENT APPLICATION NUMBER: US/10/120,988  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 09/774,528  
; PRIOR FILING DATE: 2001-01-30  
; NUMBER OF SEQ ID NOS: 441  
; SOFTWARE: pc\_ft\_genes Version 2.0

```
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205) ..(1305)
US-10-120-988-177

Alignment Scores:
Pred. No.: 9,89e-170 Length: 1416
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-120-988-177 (1-1416)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 502 ATGGAGGCTCATGCACTTACTATTAGTTGACCAATTAAGTAAGATTCTTTTATGTT 561
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 562 CTTTGTGATTTCTGGAGGTCACCTGCTGTGGCATTTAGTTCAAGAGGTTTCAGATTTCG 621
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 622 CTTCTGGAAAGTCTTTGACATAGCACCAAGTGACATGCTTGAACAAGTGGCAAGAAGA 681
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
DB 682 CTTTCTTTAAATMAAACATCCAGAGTCTCCACCAAGATGGTGGGAAAGCCATAGAACAT 741
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 742 TTGGCCAAACAGAAATAGATTTCATTGACATCAAACTCCCTTGCAATAGCTATAA 801
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 802 AATTGTGATTTTCTTTACTGGACTTCACACCGTTACTGATTAATAATATGAAAAAG 861
QY 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
DB 862 GAAAAAGGAGGATATTGAGAAAGGGCAAACTCTGCTTCAGCAGCAGACATTCCTGC 921
QY 141 ThrValGlnHisThrMetAlaCysHisAlaLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 922 ACAGTACGACACACATGGCATGTCATCTGTGAAAAAGAACACATCGGGCTATTCGTT 981
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 982 TGTAAAGCAGAGAGCTGTACTCTCAAAATATGCAAGTACTGCTGTGTGTGTC 1041
QY 181 AlaSerAsnPheIleIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
DB 1042 GCAGATGAATCTTATATCCGAGAGCTCTGGAATATTTTAAACAAACAGACAGGCACT 1101
QY 201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleThrAsnGly 220
DB 1102 TTGTTGTGCTCTCCCTCCAGACTATGACCTGATTAATGGCATTTATATGCAATGAGT 1161
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
DB 1162 ATTGAAGACACTACGCTGGCTGGCATTTTACATGACATAGAGGACATCGGCTATGAA 1221
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlyValGlyGluAlaSerIleLysVal 260
DB 1222 CCAAAATGTCCTCTTGAGATGACATATCAAAAGAGTGGAGAGCTTCATTAAGTA 1281
QY 261 ProGlnLeuLysMetGluIle 267
|||||
```

```
DB 1282 CCACATTAATAATGAGATA 1302

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:
Pred. No.: 1.11e-169 Length: 1526
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-067-443-23 (1-1526)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 1 ATGGAGGCTCATGCACTTACTATTAGTTGACCAATTAAGTAAGATTCTTTTATGTT 60
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 61 CTTTGTGATTTCTGGAGGTCACCTGCTGTGGCATTTAGTTCAAGAGGTTTCAGATTTCG 120
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 121 CTTCTGGAAAGTCTTTGACATAGCACCAAGTGACATGCTTGAACAAGTGGCAAGAAGA 180
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHis 80
DB 181 CTTTCTTTAAATMAAACATCCAGAGTCTCCACCAAGATGGTGGGAAAGCCATAGAACAT 240
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 241 TTGGCCAAACAGAAATAGATTTCATTGACATCAAACTCCCTTGCAATAGCTATAA 300
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120
DB 301 AATTGTGATTTTCTTTACTGGACTTCACACCGTTACTGATTAATAATATGAAAAAG 360
QY 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
DB 361 GAAAAAGGAGGATATTGAGAAAGGGCAAACTCTGCTTCAGCAGCAGACATTCGTC 420
QY 141 ThrValGlnHisThrMetAlaCysHisAlaLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 421 TGTAAAGCAGACACATGGCATGTCATCTGTGAAAAAGAACACATCGGGCTATTCGTT 480
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 481 TGTAAAGCAGAGACTTTTACTCTCAAAATATGCAAGTACTGCTGTGTGTGTC 540
QY 181 AlaSerAsnPheIleIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
|||||
```

Db 541 GCAGTAACCTTCTATATCCGAGCTCTGGAAATTTTACAAACGACAGTGCCT 600  
Qy LeuLeuCySPProProArgLeuCyThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
Db 601 TTGTTGTGCTCTCTCCGACGATGACATGATTAATGGCATTTATATGACAGGATGCT 660  
Qy 221 ILeGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTrpGlu 240  
Db 661 ATTGAAGACGACGAGCTGCTGGGCTTTTACATGACATGAGGACATCCGCTATGAA 720  
Qy 241 ProLysCySPProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
Db 721 CCAAAATGCTCTCTTGGAGTACATATCAAAAGAGTTGGAGAGCTTCCATAAAGTA 780  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 781 CCACATTAATAAATGAGATA 801  
RESULT 3  
US-10-649-273-23  
; Sequence 23, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; CURRENT FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-649-273-23  
Alignment Scores:  
Pred. No.: 1,11e-169 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-649-273-23 (1-1526)  
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 1 ATGAGGCTCATGCACTTACTATTAAGTTGACCAATAAAGTGAATTTCTTTTAACTT 60  
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 61 CTTTGAATTTCTGAGAGGTCACTGCTGTGGCATTAAGTTCAAGAGCTTTCAGATTTTCG 120  
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 121 CTTCTTGAAAGTCTTTTGACATGACACAGGTGACATGCTTGACAGGTGCAAGAAGA 180  
Qy 61 LeuSerLeuIleLysHisAspProGluCysSerThrMetSerGlyLysValAlaIleGluHis 80  
Db 181 CTTTCTTAATAAACAATCCAGAGTCTCCACATGAGTGGGAGAAAGCCATAAGACT 240  
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisAspIleLysProProLeuHisHisAlaLys 100  
Db 241 TTGGCCAAACAAGAGAAATTAATTTTATTTGACATCAAACTCCCTTGCAATCATCTAAA 300  
Qy 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120

Db 301 AATTGGATTTTCTTTTACTGGACTTCAACGCTTACTGATTAATTAATGAAGAAAAG 360  
Qy 121 GlnLysGluGluGlyIleLeuLysGlyGlnIleLeuSerSerLalaAspIleAlaIle 140  
Db 361 GAAAAAGAGAGAGATTAAGAAGGGCAATCTGTCTTCAGACGACGACATGCTGCTCC 420  
Qy 141 ThrValGlnHisGlyThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
Db 421 ACGATACAGACACATATGCACTGTCCTTGTGAAAAGACATGCGGCTATTTCTGTTT 480  
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
Db 481 TGTAAAGCAGAGAGACTTTGATCTCAAAATTAATGACGACTGGTTCATCTGGTGTGTC 540  
Qy 181 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
Db 541 GCAGTAACCTTCTATATCCGAGCTCTGGAAATTTTACAAACGACAGTGCCT 600  
Qy 201 LeuLeuCySPProProArgLeuCyThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
Db 601 TTGTTGTGCTCTCTCCGACGATGACATGATTAATGGCATTTATGATTGCAATGGAATGCT 660  
Qy 221 ILeGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTrpGlu 240  
Db 661 ATTGAAGACGACGAGCTGCTGGGCTTTTACATGACATGAGGACATCCGCTATGAA 720  
Qy 241 ProLysCySPProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
Db 721 CCAAAATGCTCTCTTGGAGTACATATCAAAAGAGTTGGAGAGCTTCCATAAAGTA 780  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 781 CCACATTAATAAATGAGATA 801  
RESULT 4  
US-10-651-722-23  
; Sequence 23, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; CURRENT FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-651-722-23  
Alignment Scores:  
Pred. No.: 1,11e-169 Length: 1526  
Score: 1385.00 Matches: 267  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-651-722-23 (1-1526)  
Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 1 ATGAGGCTCATGCACTTACTATTAAGTTGACCAATAAAGTGAATTTCTTTTAACTT 60

```

QY 21 LeuLeu11eserGlyGlyHisCysLeuLeu11aleuVal1GlnGlyVal1SerAspPheLeu 40
| | | | |
Db 61 CTTTGGATTTCCTGGAGGTCACGTCTGTGGCATTTAGTTCAAGAGTTTCAATTTCTG 120
QY 41 LeuLeuGly1ySerLeuAsp11eal1aProGly1aAspMetLeuAspLysVal11aArgArg 60
| | | | |
Db 121 CTTCTGGAAAGTCTTTGGACATGACACAGGTGACAGCTTGACAGAGTGGCAAGAGA 180
QY 61 LeuSerLeu11eLysHis1aProGlyCysSerThrMetSerGly1yVal1a11eGlyHis 80
| | | | |
Db 181 CTTTCTTTAATAAACAATCCAGAGTGTCCACCATGAGTGGTGGAAAGCCATTAAGACT 240
QY 81 Leu11aLysGlnGly1aAsnArgPheHis1aPheAsp11eLysProProLeuHis1a1aLys 100
| | | | |
Db 241 TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATATAGCTAA 300
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisVal1ThrAspLys11e11eMetLysLys 120
| | | | |
Db 301 AATTGTGATTTTCTTTTACCTGACCTTCAACACGTTACTGATTAATAATGAAGAAAG 360
QY 121 GlnLysGlnGly1y11eGlnLysGly1y11eLeuSerSer1a1a1aAsp11eal1a1a 140
| | | | |
Db 361 GAAAAAGAGAGAGATTAAGAGAGGAGCAAACTGTCTTCAGCAGCAGACATTTCTGCC 420
QY 141 ThrVal1GlnHisThrMetAlaCysHisLeuVal1ySarGThrHis1aArgAla11eLeuPhe 160
| | | | |
Db 421 ACACTACAGCACAACAATGCGATGTCATCTTGGAAGAAACAACAATCGGCTTATCTGTTT 480
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal1LeuVal11aSerGly1yVal 180
| | | | |
Db 481 TGTAAAGCAGAGAGCTTGTACTCTCAAAATATATGACATGCTGTTGACTCTGTGTGTC 540
QY 181 AlaSerAsnPheThr11eArgArgAlaLeuGln11eLeuThrAsnAla1aThrGlnCysThr 200
| | | | |
Db 541 GCAAGTAACCTCTATATCCGAGAGCTCTGAAATTTTAAACAAGCACAACAGTGCCT 600
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGly11eMet11eal1a1aArgArg 220
| | | | |
Db 601 TTGTTGTCTCTCTCCACGACTATGACTGATTAATGGCATTATGATTCAGATGGAATGCT 660
QY 221 11eGlnArgLeuArgAlaGly1yLeuGly11eLeuHisAsp11eGlnGly11eArgGlyArg 240
| | | | |
Db 661 ATTGAAGACTACAGCTGCTGGCTGGCATTTTACATGACATAGAGGATCCGCTATGAA 720
QY 241 ProLysCysProLeuGlyVal1aAsp11eSerLysGlnVal1GlyGlnAlaSer11eLysVal 260
| | | | |
Db 721 CCAAAATGTCTCTTGGATGAGACATATCAAAAGAAAGTTGGAAGCTTCCATTAAGATA 780
QY 261 ProGlnLeuLysMetGln11e 267
| | | | |
Db 781 CCACAAATTAAATAATGAGATA 801

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
US-10-067-443-1
Alignment Scores:
Pred. No.: 1,91e-169 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0
US-10-649-273-2_copy_148_414 (1-267) x US-10-067-443-1 (1-2197)
QY 1 MetGlnAlaHisAlaLeuThr11eArgLeuThrAsnLysVal1GlnPheProPheLeuVal 20
| | | | |
Db 672 ATGAGAGGCTCATGACATTAATTAAGTTAGTTGACCAATTAAGTGAATTTCTTTTATGAT 731
QY 21 LeuLeu11eserGlyGlyHisCysLeuLeu11aleuVal1GlnGlyVal1SerAspPheLeu 40
| | | | |
Db 732 CTTTGGATTTCCTGGAGGTCACGTCTGTGGCATTTAGTTCAAGAGTTTCAATTTCTG 791
QY 41 LeuLeuGly1ySerLeuAsp11eal1aProGly1aAspMetLeuAspLysVal11aArgArg 60
| | | | |
Db 792 CTTCTGGAAAGTCTTTGGACATGACACAGGTGACATGCTTGACAGAGTGGCAAGAGA 851
QY 61 LeuSerLeu11eLysHis1aProGlyCysSerThrMetSerGly1yVal1a11eGlyHis 80
| | | | |
Db 852 CTTTCTTTAATAAACAATCCAGAGTGTCCACCATGAGTGGTGGAAAGCCATTAAGACT 911
QY 81 Leu11aLysGlnGly1aAsnArgPheHis1aPheAsp11eLysProProLeuHis1a1aLys 100
| | | | |
Db 912 TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAAACTCCCTTGATATAGCTAA 971
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisVal1ThrAspLys11e11eMetLysLys 120
| | | | |
Db 972 AATTGTGATTTTCTTTTACTGACTTCAACGTTACTGATTAATAATTAAGAAAG 1031
QY 121 GlnLysGlnGly1y11eGlnLysGly1y11eLeuSerSer1a1a1aAsp11eal1a1a 140
| | | | |
Db 1032 GAAAAAGAGAGATTAAGAGAGGAGCAAACTGTCTTCAGCAGCAGACATTTCTGCC 1091
QY 141 ThrVal1GlnHisThrMetAlaCysHisLeuVal1ySarGThrHis1aArgAla11eLeuPhe 160
| | | | |
Db 1092 ACAGTACAGCACAACAATGCGATGTCATCTTGGAAGAAACAACAATCGGCTTATCTGTTT 1151
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaVal1LeuVal11aSerGly1yVal 180
| | | | |
Db 1152 TGTAAAGCAGAGACTTGTACTCAAAATTAATGACATGCTGTTGATCTGTGTGTC 1211
QY 181 AlaSerAsnPheThr11eArgArgAlaLeuGln11eLeuThrAsnAla1aThrGlnCysThr 200
| | | | |
Db 1212 GCAAGTAACCTCTATATCCGAGAGCTCTGAAATTTTAAACAAGCACAACAGTGCCT 1271
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGly11eMet11eal1a1aArgArg 220
| | | | |
Db 1272 TTGTTGTCTCTCTCCACGACTATGACTGATTAATGGCATTATGATTCAGATGGAATGCT 1331
QY 221 11eGlnArgLeuArgAlaGly1yLeuGly11eLeuHisAsp11eGlnGly11eArgGlyArg 240
| | | | |
Db 1332 ATGGAAGACTACAGCTGCTGGCATTTTACATGACATAGAGGATCCGCTATGAA 1391
QY 241 ProLysCysProLeuGlyVal1aAsp11eSerLysGlnVal1GlyGlnAlaSer11eLysVal 260
| | | | |
Db 1392 CCAAAATGTCTCTTGGATGAGACATATCAAAAGAAAGTTGGAAGCTTCCATTAAGATA 1451
QY 261 ProGlnLeuLysMetGln11e 267
| | | | |
Db 1452 CCACAAATTAAATAATGAGATA 1472

```

RESULT 6  
US-10-649-273-1

RESULT 5  
US-10-067-443-1

```

; Sequence 1, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```
; Sequence 1, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
; US-10-649-273-1

Alignment Scores:
Pred. No.: 1,91e-169 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-649-273-1 (1-2197)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 672 ATGAGGCTCATGCACTACTACTATTAGTTGACCAATTAAGTAATTCCTTTTAGTT 731
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 732 CTTTGATTCTGGAGGTCACTGCTGTTGGCATAGTTCAAGAGTTTCAGATTTCGTG 791
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 792 CTTCTTGAAAGTCTTTGACATAGCACACAGGTGACATGCTTGACAAAGGTGCAAGAGA 851
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 852 CTTTCTTTAATAAATCATCCAGAGTCTCCACCATGAGTGGTGGAAAGCCATTAGACAT 911
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
DB 912 TTGGCCAAACAGAGAAATAGATTTCATTTTGACATCAACCTCCCTTGATCATGCTAA 971
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValIleThrAspLysIleIleMetLysLys 120
DB 972 AATTGTGATTTTCTTTTACTGAGCTTCAACACGTTACTGATTAATATATGAAAAAG 1031
QY 121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140
DB 1032 GAAAAAGAGAAAGTATTGAGAGAGGGAATCCGTCTTCAGCAGCAGCATTCGTGCC 1091
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160
DB 1092 ACAATGACGACACACATGGCATGTCTTGTGAAAAAGACACATCGGGCTATTCTGTT 1151
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180
DB 1152 TCTAAGCAGAGACCTTGTTACTCAAAAVATGACGATCTGTTGCACTGTGTGTGTC 1211
QY 181 AlaSerAsnPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
```

```
DB 1212 GCAGTAACTCTATATCCGACAGCTCTGAAATTTTAAACAAACGACACATGCACT 1271
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTPrAsnGly 220
DB 1272 TTGTTGTGCTCTCTCCAGACTATGACTGATTAATGGCAATTATGATTGATGAAATGGT 1331
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 240
DB 1332 ATTGAAAGACTAGCTGCTGGCTTGGCATTTTATCATGACATTAAGAGCATCCGCTATGA 1391
QY 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260
DB 1392 CCAAAATGTCTCTTGAGATAGACATATCAAAAGAGTTGGAGAACTTCATTAAGTA 1451
QY 261 ProGlnLeuLysMetGluIle 267
DB 1452 CCACAAATTAATAATGAGATA 1472

RESULT 7
US-10-651-722-1
; Sequence 1, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
; US-10-651-722-1

Alignment Scores:
Pred. No.: 1,91e-169 Length: 2197
Score: 1385.00 Matches: 267
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2_COPY_148_414 (1-267) x US-10-651-722-1 (1-2197)
QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB 672 ATGAGGCTCATGCACTACTACTATTAGTTGACCAATTAAGTAATTCCTTTTAGTT 731
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB 732 CTTTGATTCTGGAGGTCACTGCTGTTGGCATAGTTCAAGAGTTTCAGATTTCGTG 791
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60
DB 792 CTTCTTGAAAGTCTTTGACATAGCACACAGGTGACATGCTTGACAAAGGTGCAAGAGA 851
QY 61 LeuSerLeuIleLysHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB 852 CTTTCTTTAATAAATCATCCAGAGTCTCCACCATGAGTGGTGGAAAGCCATTAGACAT 911
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100
```

Db 912 TTGGCCAACAAGAAATAGATTTCATTTTGACATCAAACTCCCTTGACATGCTAA 971  
Qy 101 AsnCysaspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 972 AATTGTGATTTTCTTTTACTGAGACTTCACACGTTACTGATTAATATATGAAAAAG 1031  
Qy 121 GltLysGltGltGlyIleGltLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
Db 1032 GAAAAAGAGAGATGATGAGAGAGGCAATCTCTCTTCAGCAGACATGCTGCC 1091  
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
Db 1092 ACAGTACGACACACATGAGCATGTCATCTTGAAAAAGAACATCGGCTATTCTGTTT 1151  
Qy 161 CysLysGlnArgAspLeuLeuProGlnIleAsnAlaValLeuValAlaSerGlyGlyVal 180  
Db 1152 TGTAGCGAGAGAGACTTGTGATCTCAAAATATATGAGATGCTGTGACTGTGTGTC 1211  
Qy 181 AlaSerAsnPhePheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
Db 1212 GCAAGTAACTTCTATATCCGACAGACTCTGGAATTTTACAAACGCAACACATGCTACT 1271  
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220  
Db 1272 TTGTTGTCTCTCTCCAGACTATGCACTGATATATGCAATTATGATTCATGGAATGCT 1331  
Qy 221 IleGltArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGltGlyIleArgTyrGlu 240  
Db 1332 ATTAAAAACATACGCTGCTGCTGCAATTATACATGATGAAGGCATTCGCTATATGA 1391  
Qy 241 ProLysCysProLeuGlyValAlaAspIleSerLysGlyValGlyAlaAlaSerIleLysVal 260  
Db 1392 CCAAAATCTCTCTTGAGTACATATCAAAAGATTTGAGAGCTTCAATAAAGTA 1451  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 1452 CCACACATTAATAATGAGATA 1472

RESULT 8  
US-10-667-443-21  
; Sequence 21, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 21  
; LENGTH: 1387  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-667-443-21

Alignment Scores:  
Pred. No.: 2.74e-166 Length: 1387  
Score: 1358.50 Matches: 266  
Percent Similarity: 91.10% Conservative: 0  
Best Local Similarity: 91.10% Mismatches: 1  
Query Match: 98.09% Indels: 25  
DB: 14 Gaps: 1

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-667-443-21 (1-1387)  
Qy 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlyPheProPheLeuVal 20  
|||||

Db 465 ATGAGAGCTCATGACCTTACTATTAGTTGACCAATTAAGTAAGAAATTCCTTTTATGTT 524  
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 525 CTTTGTGATTTTCTGAGGTCACTGTCTGTGGCACTTATGTTCAAGAGCTTCAGATTTTCG 584  
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 585 CTTCTTGAAAAGCTCTTTGACATAGACACAGGTGACATCTTGACAAAGGTGGCAAGAGA 644  
Qy 61 LeuSerLeuIleLysHisProGlnCysSerThrMetSerGlyGlyLysAlaIleGlnHis 80  
Db 645 CTTTCTTATTAATAACATCAGAGTGTCTCACCATGATGTGTGGAAAGCCATAGAGCAT 704  
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
Db 705 TTGGCCAAACAAGAAATGATTTCAATTTTGATCAATCAACCTCCCTGCACTCAATGCTAA 764  
Qy 101 AsnCysaspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 765 AATTGTGATTTTCTTTTACTGACTTCAACAGTTACTGATTAATAATATGAAAAAG 824  
Qy 121 GltLysGltGltGlyIle-----GltLysGlyGlnIleLeuSerSerAla 126  
Db 825 GAAAAAGAGAGATATATTTCTAATTAGTAAAGTTGAACAGATAATATTCTCGATTG 884  
Qy 127 -----GltLysGlyGlnIleLeuSerSerAla 135  
Db 885 TGCCTAAAAATAGCTGCTCATTTCTGCAAGTATGAAGAGGGCAAACTCTGCTTACGA 944  
Qy 136 AlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155  
Db 945 GAGACATTCCTGCCACAGTACACACACAAATGCAATGTCATTTGTGAAAAACACAT 1004  
Qy 156 ArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnIleAsnAlaValLeuVal 175  
Db 1005 CCGGCTATTCTGTTTGTAAAGAGAGAGACTTGTATCTCAAAATATATGCAATGCTT 1064  
Qy 176 AlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGlnIleLeuThrAsn 195  
Db 1065 GCATCTGGTGTGTGCGCAAGTAACTTATATCCGACAGCTCTGGAATTTTAAACAAC 1124  
Qy 196 AlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMet 215  
Db 1125 GGAACACAGTGAACCTTGTGTCTCTCCACAGCTATGCACTGATTAATGSCATTATG 1184  
Qy 216 IleAlaTrpAsnGlyIleGltLysArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235  
Db 1185 ATTGATGAGATGATGATTAAGAACTAGCGTGTGGCATTTTACATGACATGAA 1244  
Qy 236 GltLysArgTyrGluProLysCysProLeuGlyValAspIleSerLysGlyValGlyGlu 255  
Db 1245 GGCATCGCTGATGAACCAAAATGTCTCTGTGAGTGAACATATCAAAAGAAATGGAGAA 1304  
Qy 256 AlaSerIleLysValProGlnLeuLysMetGluIle 267  
Db 1305 GCTTCATTAATAAGTACCAATTAATAAATGAGATA 1340

RESULT 9  
US-10-649-273-21  
; Sequence 21, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 21  
LENGTH: 1387  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-649-273-21

Alignment Scores:

Pred. No.:	2,74e-166	Length:	1387
Score:	1358.50	Matches:	266
Percent Similarity:	91.10%	Conservative:	0
Best Local Similarity:	91.10%	Mismatches:	1
Query Match:	98.09%	Indels:	25
DB:	17	Gaps:	1

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-649-273-21 (1-1387)

1 MetGluAlaHsAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
465 ATGAGGCTCATGCACTTACTATTAGTTGACCAATTAAGTGAATTTCTTTTATGTT 524  
21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
525 CTTTGATTTCTGGAGGTCACGTCTGTGGCATTAGTTCAAGAGGTTTCAATTTTCG 584  
41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
585 CTTCTTGAAAGTCTTTGGACATGACACAGGTGACACGCTTGACAGGTGGCAAGAGA 644  
61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIleGlnHis 80  
645 CTTTCTTTAAATAAACATCCAGAGTCTCCACATGAGGGGGAAGCCATGAGCAT 704  
81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
705 TTGGCCAAACAGAGAAATGATTTCACTTTGACATCAAACTCCCTTGATCATGCTAA 764  
101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
765 AATTGTGATTTTCTTTTACTGCACTTCAACACGTTACTGATTAATATGAAAAAG 824  
121 GluLysGlnGluGlyIle----- 126  
825 GAAAAGAGGAGGATATTTCTAATTAGTAAAGTTGAACAGATTAATATTCCTGATGG 884  
127 -----GluLysGlnGlnIleLeuSerSerAla 135  
885 TGCCATAAAATAGCTGCTCATTTCTGCAGGTATGAGGAGGCGCAATCTGTCTTCAACA 944  
136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155  
945 GCACAGCATTTGCTGCACAGTACAGACACAATGGCATGTCATCTGTGAAAAAGACACT 1004  
156 ArgAlaIleLeuPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuVal 175  
1005 CGGGCTATTCTGTTTGTAGCAGAGACTGTGTAACCTCAAAATATATGCACTACGTGGTT 1064  
176 AlaSerGlyValAlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsn 195  
1065 GCATCTGGTGGTGGCAAGTAACTTCTATATCCGACAGGCTCTGGAATTTTAAACAAAC 1124  
196 AlaThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMet 215  
1125 GCACACAGTGCATCTTCTGTGCTCTCTCCAGACATATGACATGATATGCACTTATAG 1184  
216 IleAlaTrpAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGln 235  
1185 ATTCACAGCAATGGATTTGAAGAACACATGATGCTGGCATTTTACATGACATAGAA 1244  
236 GluLysLeuArgTrpGluProLysCysProLeuGlyValAlaAspIleSerLysGlnValGlyGlu 255

DB 1245 GCATCCGCTATGAAACCAAAATGCTCTTGGAGTACATATCAAAAGATTGGAGA 1304  
QY 256 AlaSerIleLysValProGlnLeuLysMetGluIle 267  
DB 1305 GCTTCATTAAGATACCAATTTAAAAATGAGATA 1340

RESULT 10  
US-10-651-722-21

Sequence 21, Application US/10651722

Publication No. US20040048302A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1

FILE REFERENCE: D0073 DIV

CURRENT APPLICATION NUMBER: US/10/651,722

CURRENT FILING DATE: 2003-08-29

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 10/067,443

PRIOR FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.2

SEQ ID NO 21

LENGTH: 1387

TYPE: DNA

ORGANISM: homo sapiens

US-10-651-722-21

Alignment Scores:

Pred. No.:	2,74e-166	Length:	1387
Score:	1358.50	Matches:	266
Percent Similarity:	91.10%	Conservative:	0
Best Local Similarity:	91.10%	Mismatches:	1
Query Match:	98.09%	Indels:	25
DB:	17	Gaps:	1

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-651-722-21 (1-1387)

1 MetGluAlaHsAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
465 ATGAGGCTCATGCACTTACTATTAGTTGACCAATTAAGTGAATTTCTTTTATGTT 524  
21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
525 CTTTGATTTCTGGAGGTCACGTCTGTGGCATTAGTTCAAGAGGTTTCAATTTTCG 584  
41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
585 CTTCTTGAAAGTCTTTGGACATGACACAGGTGACACGCTTGACAGGTGGCAAGAGA 644  
61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyLysAlaIleGlnHis 80  
645 CTTTCTTTAAATAAACATCCAGAGTCTCCACATGAGGGGGAAGCCATGAGCAT 704  
81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
705 TTGGCCAAACAGAGAAATGATTTCACTTTGACATCAAACTCCCTTGATCATGCTAA 764  
101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
765 AATTGTGATTTTCTTTTACTGCACTTCAACACGTTACTGATTAATATGAAAAAG 824  
121 GluLysGlnGluGlyIle----- 126  
825 GAAAAGAGGAGGATATTTCTAATTAGTAAAGTTGAACAGATTAATATTCCTGATGG 884  
127 -----GluLysGlnGlnIleLeuSerSerAla 135  
885 TGCCATAAAATAGCTGCTCATTTCTGCAGGTATGAGGAGGCGCAATCTGTCTTCAACA 944  
136 AlaAspIleAlaIleThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHis 155



Db 945 GCAGCATGCTGCACAGTACAGACCAATGCGATGCTTGTGAAAAGACACAT 1004  
Qy 156 ArgAlaIleLeuPheCysGlyLeuGlnArgAspLeuProGlnAsnAspAlaIleVal 175  
Db 1005 CGGGCTATTCCTGTTTGTAGAGAGAGACTTGTACCAAAATTAAGCACTACTGCTT 1064  
Qy 176 AlaSerGlyGlyValAlaSerAspPheTyrIleArgAlaLeuGlnIleLeuThrAsn 195  
Db 1065 GCATGTGTGTGTGCGAAGTACTTCTATTCGCGAGCTCTGGAATTTTAACAAAC 1124  
Qy 196 AlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsnGlyIleMet 215  
Db 1125 GCACACAGTGCACCTTGTGTGTCTCTCCCAAGCATATGACATGATATGCAATTAG 1184  
Qy 216 IleAlaThrPheAsnGlyIleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlu 235  
Db 1185 ATTCGATGAAATGTGATTAAGAAAGACTACGTGTGCTTGGCAATTTTACATGACATAGAA 1244  
Qy 236 GlyIleArgTyrGlnProGlyCysProLeuGlyValAspIleSerIleGlyValGlyGlu 255  
Db 1245 GGCATCCGCTATGAAACCAAAATGCTCTTGGAGTACATATCAAAAGAAAGTTGAGAA 1304  
Qy 256 AlaSerIleLeuValProGlnLeuLeuYsmetGluIle 267  
Db 1305 GCTTCATTAAGATACCAATTAATAAGAGATA 1340

## RESULT 11

US-10-012-140-6  
Sequence 6, Application US/10012140  
Publication No. US20030009017A1  
GENERAL INFORMATION:  
APPLICANT: Leiby, Kevin R.  
APPLICANT: Kapeller-Libermann, Rosana  
TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
FILE REFERENCE: THERBOP  
CURRENT APPLICATION NUMBER: US/10/012,140  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: 60/246,768  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,772  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,185  
PRIOR FILING DATE: 2000-11-15  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 1245  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-012-140-6

## Alignment Scores:

Pred. No.: 5.45e-165 Length: 1245  
Score: 1348.00 Matches: 260  
Percent Similarity: 98.50% Conservative: 3  
Best Local Similarity: 97.38% Mismatches: 4  
Query Match: 97.33% Indels: 0  
DB: 14 Gaps: 0

US-10-649-273-2\_COPY 148\_414 (1-267) x US-10-012-140-6 (1-1245)

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
Db 442 ATGAGGCTCATGCACTTACATTTAGGTGACCAATAAGTGAATTTCTTTTATGTT 501  
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 502 CTTTGTATTCGAGAGTCACTGTCTGTTCGATTAAGTTCAAGAGATTTCATTTTCG 561

Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 562 CTTCTTGAAGACTCTTTGACATAGCACAGGAGCATCTTGACAAAGGTGGCAAGAGA 621  
Qy 61 LeuSerLeuIleLysHisProGlnCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
Db 622 CTTTCTTTAATAAACATCCAGAGTCTCCACACATGAGGTGGGAAACCATAGAAACT 681  
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
Db 682 TTGGCCAAACAAAGAAATGATTTCAATTTGACATCAACCTCCCTTGACATAGCTTAA 741  
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 742 AATGTGATTTTCTTTTACTGACCTTCAACGGTACTAGATTAATAATAGAAACAG 801  
Qy 121 GlyLysGlnGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
Db 802 AAACAAGAGAAAGTATAGAAAGGGCAAACTCTGTTCAGCAGCAGCATTCCTGCC 861  
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
Db 862 ACAGTACAGCACACATGCGATGTCATCTTGAAAAAGAACATGCGGCTATTCGTGT 921  
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValIleValAlaSerGlyGlyVal 180  
Db 922 TGTAAACAGAGAGACTTGTACTCTCAAAATTAATGCAATGCTGTTGATCTGCTGTC 981  
Qy 181 AlaSerAspPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
Db 982 GCAAGTAACTCTTAATATCCGACAGAGCTGGAATTTTAACAAACGACACAGTGCAC 1041  
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrPheAsn 220  
Db 1042 TTGTTGTGCTCTCTCCACAGTATGCACTGATTAATGCAATTAATGATGATGGAATGG 1101  
Qy 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGlu 240  
Db 1102 ATTGAAAGACTACGTCCTGGCTTGGCACTTTTACATGACATAGAAAGCATCCGCTATGA 1161  
Qy 241 ProLysCysProLeuGlyValAspIleSerIleGlyValGlnAlaSerIleLysVal 260  
Db 1162 CCAAAATGCTCTTGAAGTACATATCAAAAGAAAGTTGAGAAAGCTTCATTAAGAATA 1221  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 1222 CCACATTAATAATGAGATA 1242

## RESULT 12

US-10-012-140-4  
Sequence 4, Application US/10012140  
Publication No. US20030009017A1  
GENERAL INFORMATION:  
APPLICANT: Leiby, Kevin R.  
APPLICANT: Kapeller-Libermann, Rosana  
TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND  
TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES  
FILE REFERENCE: THERBOP  
CURRENT APPLICATION NUMBER: US/10/012,140  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: 60/246,768  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,772  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,185  
PRIOR FILING DATE: 2000-11-15  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 1820  
TYPE: DNA

ORGANISM: Homo sapiens  
FEATURES:  
NAME/KEY: CDS  
LOCATION: (146)...(1390)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1820)  
OTHER INFORMATION: n = A,T,C or G  
US-10-012-140-4

Alignment Scores:  
Pred. No.: 9,61e-165 Length: 1820  
Score: 1348.00 Matches: 260  
Percent Similarity: 98.50% Conservative: 3  
Best Local Similarity: 97.38% Mismatches: 4  
Query Match: 97.33% Indels: 0  
DB: 14 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-012-140-4 (1-1820)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
DB 587 ATGAGGCTCAGTACCTTACTTAAAGTTAGCCAAATAAGTAAGATTCTTTTAACTT 646  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 647 CTTTGAATTTCTGAGGCTCAGTCTGTGGCATTAGTTCAAGAGGTTTCAGATTCTTG 706  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 707 CTTCTGGAAGTCTTTGACATAGCACAGGTGACATGCTTGAACAAGTGGCAAGAAGA 766  
QY 61 LeuSerLeuIleLysHisAspProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
DB 767 CTTTCTTAATAAACAATCCAGAGTGTCCACCATGAGTGTGGGAAAGCCATTAGACAT 826  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLys 100  
DB 827 TTGGCCAAACAGGAATAGATTCAATTTTGAATCAACCTCCCTGCATCATGCTAAA 886  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysHis 120  
DB 887 AATTGATATTTTCTTTTACTGACCTTCAACACGTTCTGATTAATAAATAAGAAACAG 946  
QY 121 GluLysGlnGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla 140  
DB 947 AAACAAAGAGAGGATATTGAGAGAGGCGAAATCTGCTTCAGCAGACATTGCTGCC 1006  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 1007 ACGATACAGCACAAATGCGATGTCATCTTGTGAAMAACAACATCGGGCTATTCTGTTT 1066  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAla 180  
DB 1067 TGTAAAGAGAGAGCTTGTACTTCAAAATATATCAGATAGCTGTTGCATCTGGTGCTTC 1126  
QY 181 AlSerAsnAspPheTyrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
DB 1127 GCAAGTAACTTCTATATCCGACAGCTCTGGAATTTTAACAAACGCAACAGTGCACT 1186  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleAsnGly 220  
DB 1187 TTGTGTGCTCTCTCCCAAGCTATGACATGATATATGCAATTAATGATGATGAGATGAT 1246  
QY 221 IleGlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGln 240  
DB 1247 ATTGAAGAAGCTACAGCTGCTGGCTTGGCATTTTAAATGACATAGAGGATCCGCTATGAA 1306  
QY 241 ProLysCysProLeuGlyValAspIleSerLysGlnValGlyGlnAlaSerIleLysVal 260  
DB 1307 CCAAAATGCTCTCTGAGATAGACATATCAAAAGAGATTGAGAAAGCTTCCATAAAGTA 1366  
QY 261 ProGlnLeuLysMetGlnIle 267

DB 1367 CCACATTAATAATGAGATA 1387

RESULT 13

US-10-094-749-400

Sequence 400, Application US/10094749

Publication No. US20030219741A1

GENERAL INFORMATION:

APPLICANT: ISOGAI, TAKAO

APPLICANT: SUGIYAMA, TOMOYASU

APPLICANT: OTSUKI, TETSUJI

APPLICANT: WAKAMATSU, AI

APPLICANT: SATO, HIROYUKI

APPLICANT: ISHII, SHIZUKO

APPLICANT: YAMAMOTO, JUN-ICHI

APPLICANT: ISONO, YUUKO

APPLICANT: HIO, YURI

APPLICANT: OTSUKA, KAORU

APPLICANT: NAGAI, KEIICHI

APPLICANT: IRIE, RYOTARO

APPLICANT: TAMECHIKA, ICHIRO

APPLICANT: SEKI, NAOHITO

APPLICANT: YOSHIKAWA, TSUTOMU

APPLICANT: OTSUKA, MOTOYUKI

APPLICANT: NAGAHARA, KENJI

APPLICANT: MASUHO, YASUHIKO

TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA

FILE REFERENCE: 08435/0160

CURRENT APPLICATION NUMBER: US/10/094, 749

CURRENT FILING DATE: 2002-03-12

PRIOR APPLICATION NUMBER: 60/350, 435

PRIOR FILING DATE: 2002-01-24

PRIOR APPLICATION NUMBER: JP 2001-328381

PRIOR FILING DATE: 2001-09-14

NUMBER OF SEQ ID NOS: 3381

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 400

LENGTH: 2208

TYPE: DNA

ORGANISM: Homo sapiens

US-10-094-749-400

Alignment Scores:

Pred. No.: 7,67e-146

Score: 1204.00

Percent Similarity: 89.51%

Best Local Similarity: 86.93%

Query Match: 17

DB: 1

Length: 2208  
Matches: 239  
Conservative: 0  
Mismatches: 4  
Indels: 24  
Gaps: 1

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-094-749-400 (1-2208)

QY 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
DB 765 ATGAGGCTCAGTACCTTACTTAAAGTTAGCCAAATAAGTAAGATTCTTTTAACTT 844  
QY 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 845 CTTTGAATTTCTGAGGCTCAGTCTGTGGCATTAGTTCAAGAGGTTTCAGATTCTTG 904  
QY 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
DB 905 CTTCTGGAAGTCTTTGACATAGCACAGGTGACATGCTTGAACAAGTGGCAAGAGA 964  
QY 61 LeuSerLeuIleLysHisAspProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
DB 965 CTTCTTAATAAACAATCCAGAGTGTCCACATAGAGTGGGAAAGCCATAGAACT 1024  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisAlaLys 100  
DB 1025 TTGGCCAAACAGGAATAGATTCAATTTTGAATCAACCTCCCTGCATCATGCTAAA 1084  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120

|||||  
Db 1085 AATTGATGATTTCTCTTACTGGAAGCTTCAACAGCTTACGTAATAATATATGAAAAAG 1144  
Qy 121 G|U|U|S|G|U|G|U|G|Y|I|E|G|U|U|S|G|Y|G|I|N|I|E|U|S|E|S|E|R|A|A|A|S|P|I|E|A|A|A| 140  
Db 1145 GAAAAAGAGGAAAGGATTTGAGAAAGGGGCAAAATCTGTCTTCAGCAGACATTCCTGCC 1204  
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleLeuPhe 160  
Db 1205 ACAATTACGACACACAAATGGCAGATGTCATCTTGTGAAAGAACACATCGGGCTATTCGT 1264  
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180  
Db 1265 TGTAAAGCAGAGAGACTTGTACCTCAAAATATGACAGTACTGTTGCATCTGTGTGTC 1324  
Qy 181 AlaseranpheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
Db 1325 GCAAGTAACTTCTGTATCCGACGAGCTCTGCAAAATTTTAAACAAACGCAACAGTGCAC 1384  
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220  
Db 1385 TTGTTGTGTCCTCTCCAGACTATGACATGATTAATGGCATTATGATTGCA----- 1435  
Qy 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrglu 240  
Db 1435 ----- 1435  
Qy 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
Db 1436 ---TGATGTCTCTCTTGAGTGAACATATCAAAAGAGTTGGAGAGCTTCCTCAATAAGTA 1492  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 1493 CCACAATTAAAAATGAGAGTA 1513

## RESULT 14

US-10-723-860-7447  
; Sequence 7447, Application US/10723860  
; Publication No. US2004025306A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Nataasha  
; APPLICANT: Ginsburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
; FILE REFERENCE: 05882.0193.NPUS01  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: US/10/723,860  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 7447  
; LENGTH: 2890  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (646)..(657)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-723-860-7447

## Alignment Scores:

Pred. No.: 1,15e-145 Length: 2890  
Score: 1204.00 Matches: 239  
Percent Similarity: 89.51% Conservative: 0  
Best Local Similarity: 89.51% Mismatches: 4  
Query Match: 86.93% Indels: 24  
DB: 18 Gaps: 1

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-723-860-7447 (1-2890)

Qy 1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20

|||||  
Db 1442 ATGAGGCGTCATGCACTTAACTATAGTTGACCAATTAAGTGAATTCCTTTTAGTT 1501  
Qy 21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspLeu 40  
Db 1502 CTTTGGATTTCTGAGGTCACATGCTGTGGCATAGTTCAAGAGATTTTCAGATTTTCG 1561  
Qy 41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
Db 1562 CTTCTTGAAAGTCTTTGGACATAGCACAGGATGACATCTTGACAAAGGTGGCAAGAGA 1621  
Qy 61 LeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIleGluHis 80  
Db 1622 CTTCTTTAATAAATCATCAGAGTCTCCACATAGTGGTGGGAACCATAGAAACT 1681  
Qy 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
Db 1682 TTGGCCAAACAAAGAAATGATTTCAATTTGACATCAAACTCTCTTGACATCATGCTAAA 1741  
Qy 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
Db 1742 AATTGATTTTCTCTTTACTGACTCAACAGTTACTGATTAATAATATGAAAAAG 1801  
Qy 121 G|U|U|S|G|U|G|U|G|Y|I|E|G|U|U|S|G|Y|G|I|N|I|E|U|S|E|S|E|R|A|A|A|S|P|I|E|A|A|A| 140  
Db 1802 GAAAAAGAGGAAAGGATTTGAGAAAGGGGCAAAATCTGTCTTCAGCAGACATTCGTGCC 1861  
Qy 141 ThrValGlnHisThrMetAlaCysHisLeuValIysArgThrHisArgAlaIleLeuPhe 160  
Db 1862 ACAGTACAGCACACATAGTCATCTTGAAAGAAACATCGGGCTAATTCGTGTT 1921  
Qy 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyVal 180  
Db 1922 TGTAAAGCAGAGAGACTTGTACCTCAAAATATGACAGTACTGTCATCTGTGTGTC 1981  
Qy 181 AlaseranpheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200  
Db 1982 GCAAGTAACTTCTGTATCCGACGAGCTCTGCAAAATTTTAAACAAACGCAACAGTGCAC 2041  
Qy 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaThrAsnGly 220  
Db 2042 TTGTTGTGTCCTCTCCAGACTATGACATGATTAATGCAATTATGATTGCA----- 2092  
Qy 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyrglu 240  
Db 2092 ----- 2092  
Qy 241 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 260  
Db 2093 ---TGATGTCTCTCTTGAGTGAACATATCAAAAGAGTTGGAGAGCTTCCTCAATAAGTA 2149  
Qy 261 ProGlnLeuLysMetGluIle 267  
Db 2150 CCACAATTAAAAATGAGAGTA 2170

## RESULT 15

US-10-067-443-20  
; Sequence 20, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D00073 NP  
; CURRENT FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 60/266,518  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 20

LENGTH: 14364  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-067-443-20

Alignment Scores:

Pred. No.:	3,84e-95	Length:	14364
Score:	824.50	Matches:	240
Percent Similarity:	26.73%	Conservative:	0
Best Local Similarity:	26.73%	Mismatches:	2
Query Match:	59.53%	Indels:	657
DB:	14	Gaps:	4

US-10-649-273-2\_COPY\_148\_414 (1-267) x US-10-067-443-20 (1-14364)

```
QY      1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20
DB      10845 ATGAGAGCTCATGACCTTACCTTATAGGTGACCAATAAAGTAGAATTTCTTTTATAGTT 10904

QY      21 LeuLeuLysSerGlyLysHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB      10905 CTTTGTATTTCTGGAGGTCACCTGCTGCTGTTGGCATTTAGTTCAAGAGGTTTCAGATTTTCTG 10964

QY      41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysVal----- 57
DB      10965 CTTTCATGAAAAGCTTTTGGACATAGCACAGGTGACATGCTTGACAAAGT-AAATTAGAA 11023

QY      57 ----- 57
DB      11024 TTAATTTCTCCATTCCTTTTTGTATGTGTGCATTTCACTAAAGTAGCAATAGATGTGC 11083

QY      57 ----- 57
DB      11084 TACCACCATTCCTAAATATTTCTGATTTTATCTTAGTAACTGAAAAAATTCACAT 11143

QY      57 ----- 57
DB      11144 ATGTGAGAAAAAATAGAAAAGTAGTACACAAATTTTATATCTTAGCCTTTCTTAATA 11203

QY      57 ----- 57
DB      11204 AAATGTAGAGGTTCAATCTGTACATTAAGGCTGAAATAGTTTGACATACAGTTATG 11263

QY      57 ----- 57
DB      11264 TAATTTGCCAAATATATGTATGTAGAAAGACGTGCTTGTAACCTAACACTGCAAAAAA 11323

QY      57 ----- 57
DB      11324 GGTAAATAAGAGATATATATAGATTAACTAAGAGACATTAAAGATGCATGACAGAA 11383

QY      57 ----- 57
DB      11384 TTAATACACAAATTAATTACTTACACACAGACAGGTCCTCCCAACCTCCCTTTGTTTAG 11443

QY      57 ----- 57
DB      11444 AATACTACAGAGGCTACATCATATATAGAAAACAAACAAACAAACAAACACTG 11503

QY      57 ----- 57
DB      11504 CTTCCACAGCTGAATAATATAGAAAGTATAGCAAGTTCTTAATATAGCGTTCACTCAT 11563

QY      57 ----- 57
DB      11564 AAGCATTTATTTGCACTTCAAGCCCAATTTTCCAAACCAATAGAAAGAGAAACATAGACAG 11623

QY      57 ----- 57
DB      11624 GGGCAGATTTGGCCCTTATTTATTTGGGTGCATATAGAAACAGGGTGTCTGCTTACC 11683

QY      57 ----- 57
```

```
DB      11684 TGAATATCAGCTATATAGTCTATATATTTGCCAAAGTAGATAGCTTTTATTCATTCAGGGGGTT 11743
QY      57 ----- 57
DB      11744 TTTTGTTTTGTAGTAATTTTCAATTTATTTCTTTGACCTTTTGGTTTACACATATTT 11803

QY      58 ----- 11804 AATTTTATGACTTAATAAATATATGTTCTTTTGTATAGTGGCAAGAGACCTTTCTTTAATA 11863

QY      65 LysHisProGlyCysSerThrMetSerGlyLysValAlaGlnHisLeuValLysGln 84
DB      11864 AAACATCCAGAGTGTCTCCACCATGAGTGGGAAAGCCATAGAAACATTTGGCCAAACAA 11923

QY      85 GlyAsnAspPheHisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPhe 104
DB      11924 GGAATATGATTTTCATTTTGAATCAACAACTCCCTTGATCATGCTTAAATTTGTGATTTT 11983

QY      105 SerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlyLys 124
DB      11984 TCTTTTACTGACCTTCAACACCTTACTGATTAATAATATATGAAAAAGAAAAAGAGAA 12043

QY      124 ----- 124
DB      12044 GGTATATTTCTAATTAGTAAAGTTGAAACAGATTAATATTTCTGTGATTTGCTTAAATA 12103

QY      125 -----GlyTlleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAl 139
DB      12104 GCTGCTCATTTCTGCAAGTATTTAGAAAGGGGCAAACTCCTTGATCATGCTTAAATTTGTGATTTT 12163

QY      139 AlaIleThrValGlnHisSerThrMetAlaCysHisLeuValLysAspThrHisArgAlaIleLe 159
DB      12164 TGCCACAGTACGACACACACATGCAATGCAATGCTTTGTGAAAAAGAACACATGGGCTATTTCT 12223

QY      159 uPheCysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeu----- 174
DB      12224 GTTTTGTAGCAGAGAGACTTGTACTTCAAAATATATGACATGACTGTAAAGTTTATCTC 12283

QY      174 ----- 174
DB      12284 AATTTATATGTAATATATGTTACACTTTCGCAATATGTATCTTTTTCCTCAAGACCTTGACCTTG 12343

QY      174 ----- 174
DB      12344 TGTTTAGATGAAACAGATCTTATATGCTTATATGCTTACCTGACAGTATGAATTTATGAG 12403

QY      174 ----- 174
DB      12404 GATAGAAAAGACTAACAGCATTTCTTGATAGTTTGTAGCTTTATAGGACAGCTGTA 12463

QY      174 ----- 174
DB      12464 TAGCTTTATGACATATAGTCTAAATTTTGCATCTTCTGTGATTTTAAAGAGGGCTT 12523

QY      174 ----- 174
DB      12524 ACAATTAAGAAAGTAATGACAGTAACTGCTATCACTATTTTAAAGAAAAATAGGTGATTT 12583

QY      174 ----- 174
DB      12584 CCTTATCTTTGATGAATAATCCCTTTGTTTGTGTTTATTAATAGCAGTCAATTT 12643

QY      174 ----- 174
DB      12644 AGCAGTGGAGGTGTATTTCAACTTTGTGACACTAATGTTGATTAAGTTCTGATATATC 12703

QY      174 ----- 174
DB      12704 CACTATATTTTACACACCAAAATCCCTTATATTTGTGCTTTAAAGCTTGAACAAACATCTC 12763

QY      174 ----- 174
DB      12764 GTTAACTGTATCTTAACTTATTTATTTAAATAATTAATTAACCTAAAGTGGGAAATGTT 12823
```

```
QY      174 ----- 174
Db      12824 TAAATGTAAGTAATTCATAGATGGAATTTTACATGATATCAAGAAATATTTTTCAGA 12883
QY      174 ----- 174
Db      12884 GTTATGTAGTAAATGCAAAATATATAAAATTTCAAGGTCTAAATAGTACTATAGA 12943
QY      174 ----- 174
Db      12944 TTGAAATTAATATAAATAATATTAGATGAAGGTGGAAGAAATATACAAAAATGCT 13003
QY      174 ----- 174
Db      13004 AGTAATGTTGTATGCTATTAGAAATTAATTAATTTTCTTCCAAATTTTATTATAC 13063
QY      174 ----- 174
Db      13064 ATGATATATGTCATCTGCCATTTACCATCTCAAAATGGATAGTTATTATTGTTAANG 13123
QY      174 ----- 174
Db      13124 CTGATATTTTCTCCAGGTTAAATTAATGACGCTGGTTCATATCCATATATATGATTAAT 13183
QY      175 -----ValAlaSerGlyGlyValAlaSerAspPheTyrIleAr 187
Db      13184 TTGGTTTTCATCAATTCCTTCAGGTTCATCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 13243
QY      187 GATGAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAr 207
Db      13244 CAGAGCTCTGGAATTTTAAACAAACGCAACAGTGCATTTGTTGTGTCTCTCCCGAG 13303
QY      207 gLeuCysThrAspAsnGlyIleMetIleIleAr 218
Db      13304 ACTATGCACTGATATGTCATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 13363
QY      218 ----- 218
Db      13364 TCATTACTATGTAATAATTAATTCGCAATTTATCATTAAGCTTCTTCCTTCAGATCT 13423
QY      219 -----AsnGlyIleGluArGLeuArGA 226
Db      13424 TGGAGCTATGATTTTATTTTAATGCTTCTTATTTAGGAATGATTTGAAAGACTTACGTG 13483
QY      226 laGlyLeuGlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLys 242
Db      13484 CTGGCTTGGGCAATTTACATGACATGAGAAGCATCCGCTATGAACCAAG 13533
```

Search completed: February 17, 2005, 01:27:09  
Job time : 559 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:38:06 / Search time 12.1891 Seconds  
(without alignments)  
2107.605 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHALTIRLTKVKEPFLV.....DISKEVGASIKVPOLKMEI 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	409.5	29.6	463	2 E84888	probable O-sialogl
2	338	24.4	365	2 AB2902	O-sialoglycoprotei
3	338	24.4	366	2 D97677	probable O-sialogl
4	338	24.4	387	2 E71711	probable O-sialogl
5	334.5	24.2	359	2 AB3274	O-sialoglycoprotei
6	329.5	23.8	344	2 E97707	O-sialoglycoprotei
7	326.5	23.6	367	2 F87257	peptidase M22 fami
8	291.5	21.0	341	2 H83572	O-sialoglycoprotei
9	290.5	21.0	335	2 G70369	O-sialoglycoprotei
10	289.5	20.9	342	2 H64074	O-sialoglycoprotei
11	279.5	20.2	337	2 A10079	probable glycoprot
12	279.5	20.2	337	2 AG0892	probable glycoprot
13	278.5	20.1	337	2 C91122	probable O-sialogl
14	278.5	20.1	337	2 B85967	probable O-sialogl
15	276	19.9	421	2 T18825	hypothetical prote
16	275.5	19.9	337	1 QOECR6	O-sialoglycoprotei
17	275	19.9	412	2 T40889	probable proteinase
18	270.5	19.5	325	2 A38108	O-sialoglycoprotei
19	265.5	19.2	354	2 C81040	O-sialoglycoprotei
20	263.5	19.0	354	2 C81986	probable O-sialogl
21	263	19.0	348	2 D82807	O-sialoglycoprotei
22	256	18.5	346	2 H70195	O-sialoglycoprotei
23	253	18.3	336	2 C97888	O-sialoglycoprotei
24	250.5	18.1	346	2 F69786	glycoprotein endop
25	250	17.7	336	2 B95015	glycoproteinase fa
26	244.5	17.7	343	2 D63718	glycoprotein endop
27	243	17.5	327	2 G72411	hypothetical prote
28	242.5	17.5	348	2 S75548	O-sialoglycoprotei
29	241.5	17.4	346	2 AF1820	O-sialoglycoprotei

30	239.5	17.3	344	2 AC1334	glycoprotein endop
31	237.5	17.1	340	2 B97011	probably O-sialogl
32	236.5	17.1	336	2 B84936	O-sialoglycoprotei
33	233	16.8	323	2 G69388	O-sialoglycoprotei
34	231	16.7	344	2 AB1705	glycoprotein endop
35	229.5	16.6	346	2 G86661	O-sialoglycoprotei
36	222.5	16.1	324	2 F75029	O-sialoglycoprotei
37	220	15.9	338	2 A71545	probable O-sialogl
38	219.5	15.8	334	2 H69056	O-sialoglycoprotei
39	219	15.8	335	2 E81278	probable glycoprot
40	218	15.7	344	2 H72106	O-sialoglycoprotei
41	218	15.7	344	2 B86515	O-sialoglycoprotei
42	217	15.7	341	2 G89996	hypothetical prote
43	216	15.6	407	2 S50740	OR17 protein - yea
44	215.5	15.6	324	2 C71215	O-sialoglycoprotei
45	215.5	15.6	344	2 H70737	probable O-sialogl

ALIGNMENTS

RESULT 1  
E84888  
probable O-sialoglycoprotein endopeptidase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Jun-2003  
C:Accession: E84888  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MID:20083487; PMID:10617197  
A:Accession: E84888  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-463 <STO>  
A:Cross-references: GB:AE002093; NID:g2583127; PIDN:AA82636.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g45270  
A:Map position: 2  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 29.6%; Score 409.5; DB 2; Length 463;  
Best Local Similarity 36.6%; Pred. No. 5.1e-29;  
Matches 100; Conservative 37; Mismatches 91; Indels 45; Gaps 6;

QY	1	MEAHALTIRLT-NKVEPFLVLLISGHCIALAVGVSDFLLGSKSLDIAPGMDLPKVAR	59
DB	192	MEAHALTIRLT-NKVEPFLVLLISGHCIALAVGVSDFLLGSKSLDIAPGMDLPKVAR	251
QY	60	RLSLIKHEPCTSMGSKAIEHLAKQGNRPHFDIKPLHNAKQDSFPGIJOHTDKIMK	119
DB	252	WLGDMH-----RSGGPAVEBELALSGDAKSVENPMKHKQCNFSYAGLKQVRLATA	306
QY	120	KEKEGIEKGGQTLSSAADIATVQNTMACHLYKTRTRAILFCQKQRLDQNNAVLVASG	179
DB	307	KE-----IRNRADIASFORVAVALHEKEKERAIDWLE--LSPSIKHWISGG	353
QY	180	VASNFYIRALEILNNAQCTLLCPPEPLCTNGMIANGIERLAPAGGLHIDIGIRY	239
DB	354	VASNFYIRALEILNNAQCTLLCPPEPLCTNGMIANGIERLAPAGGLHIDIGIRY	403
QY	240	B-----PKCPGLVDISKVEGSA	256
DB	404	DEPPATREPDVYDLRPRWPLGSEYAKGRSEA	436

RESULT 2  
AB2902  
O-sialoglycoprotein endopeptidase gcp [imported] - Agrobacterium tumefaciens (strain C58,  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C:Accession: AB2902  
R:Wood, D.W.; Seibubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, I.;  
erage, G.; Gillet, W.; Grat, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, B.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AB2902  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-365 <NUR>  
A:Cross-references: GB:AE008688; PIDN:AAL43632.1; PID:g17741154; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: gcp  
A:Map position: circular chromosome  
C:Superfamily: O-sialoglycoprotein endopeptidase

	Query Match	24.4%	Score 338;	DB 2;	Length 365;
	Best Local Similarity	34.9%	Pred. No. 1.1e-22;		
	Matches	90;	Conservative	40;	Mismatches 96; Indels 30; Gaps 7;
Qy	1	MEAAHTIRLTNKPVEPPFVLVLSGGHCLTALVGVSDPFLLGKSLIDAPGMDLKVARR	60		
Db	120	LEGHATLATLDGLSPFYMLTVSGGHTQLVLRVGGEYERKGTITTDLAGAEPDTALC	179		
Qy	61	ISLIKHPECSWTMSGKAIEHLAKOGNRPHFDIKPEPLHHAKNCDFSEFTGI----	QHVTDKI 116		
Db	180	LGL-PYP-----GGPAVENAAAGDPDRFPFLPRPMVGEARLDFFSGLKTAVRQAATAI	232		
Qy	117	INKKKEEKEGIIKQQILSSAADIAATVQHTMACHLVKRTHRALLFKQRDLTFQNNA--VL	174		
Db	233	APLSEDD-----IADICASFOKAVSRRLKDIRIGRGLARFVE--PFHINGEPAL	279		
Qy	175	VASGGVASNPYIRRALEILTNAQCILLCPRPRLCTDGMIMANNGIEPTLRAGLGILHDI	234		
Db	280	VVAGGVAANOELRQTLOALCDTHGFREFVAPPHRLCTDYANAAMIAMGLERMABG----	RQA 335		
Qy	235	BGRIVEPKCEIVGDISK	252		
Db	336	DALBVAAPRSRWPIJDSAE	353		

RESULT 3  
 D97677  
 Probable O-sialylglycoprotein endopeptidase (glycoproteinase) [imported] - Agrobacterium  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C/Accession: D97677  
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
 A.; Liu, F.; Molian, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.  
 Science 294, 2333-2328, 2001  
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A/Reference number: A97359; MUID:21608551; PMID:11743194  
 A/Accession: D97677  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-366 <KUR>  
 A/Cross-references: GB:AE007869; PIDN:AAK8373.1; PID:g15157858; GSPDB:GN00169  
 C/Genetics:  
 A/Gene: AGR\_C\_4806  
 A/Map position: circular chromosome  
 A/Superfamily: O-sialylglycoprotein endopeptidase

Query Match 24.4% Score 338 DB 2 Length 366;  
Best Local Similarity 34.9% Pred. No. 1.1e-22;  
Matches 90; Conservative 40; Mismatches 96; Indels 30; Gaps 7.

Oy 1 MEAAATLRLTNKVEFPPLVLLISGGHCLLALVOGVSDPFLILGSKSLDIAPGDMDLRVAR 60  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 121 LEGAATLRLLDGISFPYLTMLLVSGGHQTLVVRVGVEERMGTTILDALCEAPDKTKT 180  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY 61 LSLIHPESTMSGGKAIHLAKOGNRFPHDIKPLHLHAKNCPSPGL-----QHWTDKI 116
Db 181 LGL-FYP-----GGPVENMAAAKQDPDRFLPFRPWGGEARLDPFSGLTKTAVQAATAI 233
QY 117 IMKKEKEBEGIGQILSSADIATATVOHTMACTLVKTRTHAILFCCKRDLLPONNA--VL 174
Db 234 APLSBD-----IADIASFQKAVSRLLKORIGLARFKVE--FPHINGEPAL 280
QY 175 VASGVAASFYTRRALEILTNAOTCTLLCPPRLCTDNGIMIAMNGIERLRAGIGILHDI 234
Db 281 VVAGGVAANOEROTLOALCDTHGFRFVAPPHRLCTDNAMIAMWAGDERMAEG----RQA 336
QY 235 EGIRRYEPCPKLGVDISKE 252
Db 337 DALEVAFPRSRWFUDGSAE 354

```

RESULT 4  
 E71711  
 Probable O-sialoglycoprotein endopeptidase (gcp) RP037 - Rickettsia prowazekii  
 C1Species: Rickettsia prowazekii  
 C1Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 09-Jul-2004  
 C1Accession: E71711  
 R1Andersson, S.G.E.; ZomorrodiPour, A.; Andersson, J.O.; Sichertz-Ponten, T.; Almark, U.  
 Nature 396, 133-140, 1998  
 A1Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
 A1Reference number: A71630; MUID:99039459; PMID:9923893  
 A1Accession: E71711  
 A1Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A1Molecule type: DNA  
 A1Residues: 1-387 <AND>  
 A1Cross-references: UNIPROT:O9ZEA8; GB:AJ235270; GB:AJ35269; NID:g3860572; PIDN:CAA14508  
 A1Experimental source: strain Madrid E  
 C1Genetics:  
 A1Gene: gcp, RP037  
 C1Superfamily: O-sialoglycoprotein endopeptidase

Query Match	Similarity	24.4%	Score 338	DB 2	Length 387
Best Local	Similarity	32.5%	Pred. No. 1.1e-22		
Matches	Conservative	87	Mismatches	83	Indels 60, Gaps 6
Qy	1	MEAAATLRLTKVPEPFLVLLISGGHCLLAVQGVSDFLGLKSLDIABDMLDKVARR	60		
Db	113	LEGHALLTARLTNTNISPYLLILLASGGHCGFVAVLGGLTKYLITTDIDAVGETFDKAKM	172		
Qy	61	LSLRKHPECSITMSGGKALIEHLAKQGNRRFHPDIKPRLHAKNCDFSFTGLQHTVDKIIMK	119		
Db	173	LNL-----SPGGGEIEKRAKLGNPHKYKPKPIINSQNCMSFSGLTAVATLML	225		
Qy	120	KEKEGIEIKGQILSSADYIAATVQHTMACHLVKTRAILFCQ-----	163		
Db	226	KEVNDV-----INDIAASFQFTIGAILSSKMDAIRLYKQINDYEDINHPTKAN	277		
Qy	164	-----RD-----LGPQN-----NAVLVASGCVASNFIRRLLEILTIN	195		
Db	278	LKSPKDEENMKPLECITRKYRRIHQNSYRSNLTNDITVIAGVANKYLQETILSDCTR	337		
Qy	196	ATQCTLLCPRPRLCTDNGIMIAMNGIER	223		
Db	338	PYGRYLIAPPMELCTDNAAIMIAVAGIER	365		

RESULT 5  
 AB3274  
 O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - *Brucella melitensis* (strain  
 C)Species: *Brucella melitensis*  
 CDate: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #ext\_change 15-Feb-2002  
 CAccession: AB3274  
 R:DelVecchio, V.G.; Kaparatel, V.; Radkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.  
 ; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Leeseck  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 Affile: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*

A/Reference number: AD3252; PMID:11756688

A/Accession: AB3274

A/Status: Preliminary

A/Molecule type: DNA

A/Residues: 1359 <KIR>

A/Cross-references: GB:AE008917; PIDN:AAU51357.1; PID:G17982056; GSPDB:GN00190

A/Experimental source: strain 16M

A/Genetics:

A:Gene: BME10175

A:Map position: 1

C/Superfamily: O-sialoglycoprotein endopeptidase

C/Keywords: hydrolase; metalloproteinase

Query Match 24.2%; Score 334.5; DB 2; Length 359;

Best Local Similarity 37.6%; Pred. No. 2.2e-22;

Matches 86; Conservative 33; Mismatches 83; Indels 27; Gaps 6;

1 MEAAHLLTRLTNKKPFPVLLISGHCILALVQGVSDPFLIGSKLIDAPGMDLKVARR 60

116 LBGHALTARLTDGLPFPYLLLVSGHQTQWLVVIGYERLGTITDIALGEAFDKAKL 175

61 LSLIKPEECSTMGSGKAIIEHLAKOGNRFHDIKPPHLHAKNCDPSFTGL-----QHTDK 115

176 LGL-PYP-----GGPVERMALQGDQKRFALPRPLKGEARLDPSFGSLKTAFAVQTATEL 228

116 IIMKKEBEGIKGQILSSADIAATVOHTMACHLVKRTTHRAILFCQRDLDPQ--NNAVL 174

229 VPLTDD-----DVTDCISFOAAVADTISDVGRSLERPKTE--PPDCATPSL 274

175 VASGVASNFYIRRALBITLTAATCTLLCPPRLCTDNGIMIANNGIER 223

275 VVAGVAAANTIRALLENLCTRHFAPFIAPLNLCTDNAAIMIANNGAR 323

#### RESULT 6

E97707 O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - *Rickettsia conorii* (strain

C/Species: *Rickettsia conorii*

C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001

C/Accession: E97707

R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro

Science 293, 2093-2098, 2001

A/Title: Mechanisms of Evolution in *Rickettsia conorii* and *Rickettsia prowazekii*.

A/Reference number: A97700; MUID:21442074; PMID:11557893

A/Accession: E97707

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-344 <KUR>

A/Cross-references: GB:AE006914; PIDN:AAU02599.1; PID:G15619097; GSPDB:GN00173

A/Genetics:

A:Gene: gcp

C/Superfamily: O-sialoglycoprotein endopeptidase

C/Keywords: hydrolase; metalloproteinase

Query Match 23.8%; Score 329.5; DB 2; Length 344;

Best Local Similarity 36.0%; Pred. No. 5.8e-22;

Matches 81; Conservative 40; Mismatches 85; Indels 19; Gaps 5;

1 MEAAHLLTRLTNKKPFPVLLISGHCILALVQGVSDPFLIGSKLIDAPGMDLKVARR 60

113 LBGHALTARLTDNI PFPYLLLVSGHGOFAVAGLGRKTLGSTTIDVABEAPDKAKM 172

61 LSLIKPEECSTMGSGKAIIEHLAKOGNRFHDIKPPHLHAKNCDPSFTGLQHTVDKIIMK- 119

173 LNL-----APFGPEIEKRAKLGDPHKYKFPKPIINSGNMNSFSGLKTAFAVTLIMTL 225

120 KEKEBEGIKGQILSSADIAATVOHTMACHLVKRTTHRAILFCQ--RDLLPNNAAVLVASG 178

226 KEINDTV-----INDIAASFQITGIELISKVDIAIRAYEOTINNFDKN--IYVAG 275

179 GVASNFYIRRALBITLTAATCTLLCPPRLCTDNGIMIANNGIER 223

276 GVAAKTLQKILSSCAATYGVRLIYPIHLCTDNAAIMIANNGAR 320

#### RESULT 7

P87257 peptidase M22 family protein [imported] - *Caulobacter crescentus*

C/Species: *Caulobacter crescentus*

C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C/Accession: P87257

R.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.I

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Grimm, M.L.; Haft, D.H.; Kolon

n, J.; Kholmova, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of *Caulobacter crescentus*.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: P87257

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-367 <STO>

A/Cross-references: GB:AE005673; NID:G13421168; PIDN:AAK22058.1; GSPDB:GN00148

A/Genetics:

A:Gene: CC0071

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 23.6%; Score 326.5; DB 2; Length 367;

Best Local Similarity 36.3%; Pred. No. 1.2e-21;

Matches 89; Conservative 36; Mismatches 99; Indels 21; Gaps 6;

1 MEAAHLLTRLTNKKPFPVLLISGHCILALVQGVSDPFLIGSKLIDAPGMDLKVARR 60

124 LEGHVASRLGADIVPFPYLLLVSGHCQLLEVSQGACKRLGTITDDAAGBAFDPKAKS 183

61 LSLIKPEECSTMGSGKAIIEHLAKOGNRFHDIKPPHLHAKNCDPSFTGLQHTVDKIIMK 120

184 LGL-PYP-----GGPALKELAVGSDPTRYALPRALLGRKDDPSFGSLKTAARIAETL 236

121 EKEBEGIKGQILSSADIAATVOHTMACHLVKRTTHRAILFCQRDLDPQNNAAVLVASGV 180

237 TTDD-----ARRDLAAGVQAALIAQLSRVDRAMTLK--DSHDEBDLRFVAVAGV 285

181 ASNFYIRRALBITLTAATCTLLCPPRLCTDNGIMIANNGIERLRAGLIGIDEGIRYE 240

286 AANGAVRAALADCEKNGFSFAAPFLAVCTDNAAIMIANNGARL--ALGIFPDLDLIA-R 342

241 PKCPL 245

343 PRWPL 347

#### RESULT 8

H83572 O-sialoglycoprotein endopeptidase PA0580 [imported] - *Pseudomonas aeruginosa* (strain PA0)

C/Species: *Pseudomonas aeruginosa*

C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C/Accession: H83572

R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bri

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kaas, A.; Laidig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathog

A/Reference number: A82950; MUID:20437373; PMID:10984043

A/Accession: H83572

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-341 <STO>

A/Cross-references: GB:AE004494; GB:AE004091; NID:G9946446; PIDN:AA03969.1; GSPDB:GN0013

A/Experimental source: strain PA01

A/Genetics:

A:Gene: gcp

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 21.0%; Score 291.5; DB 2; Length 341;

Best Local Similarity 35.4%; Pred. No. 1.5e-18;

Matches 87; Conservative 37; Mismatches 99; Indels 23; Gaps 9;



[illegible]

Query	Match	21.0%	Score 290.5	DB 2	Length 335
Db	112	LEGIYVFEKKYEPPLALII	SGGHTDLVDPGRYDPLGGTLD	AVEAYDKAKM	171
Qy	61	LSLIIKHECSTMSGKAIE	HLAKQNRPHFDIKPLHAKKNC	DFSFTGJQHVTDKIIMKK	120
Db	172	LGL-GYP-----	GGPIIDRAKGGKCL-YLP	PRPMEGNTANSFSLK---	TAIINTLL 220
Qy	121	EKEEGIEKGOILLSSAD	IAATVQHTMACHLVKRTHRA	ILFCQKORDLLFPONNAV	LVASGV 180
Db	221	KKEKQVRK-----	EDIAVSFOETVAILEKS---	LMMKKTGKIR----	LVVGVG 265
Qy	181	ASNYIRALEIILTNQ---	CTLCPPRCLCTNGMIAMNG	IERLAPAGGILHDIEGI	237
Db	266	SANSRLR--EVRKASQ	EGFELYIPHPISLTNDAL	IAAYAGMERFRGGVAP	LDVNP- 321
Qy	238	RYEKCP 245			
Db	322	--QENIP 327			

RESULT 10

H64074

O-sialoglycoprotein endopeptidase (BC 3.4.24.57) - Haemophilus influenzae (strain Rd KW2)

N-alternate names: sialoglycoproteinase

C:Species: Haemophilus influenzae

C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004

C/Accession: G70369

R/Dickert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lennox, A.L.; Graham, D.E.; O'V.

Nature 392, 353-358, 1998

A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A/Reference number: A70300; MUID:98196666; PMID:9537320

A/Accession: G70369

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1535 <AOP>

A/Cross-references: UNIPROT:O66986; GB:AE000708; NID:g2983356; PIDN:AAC06951.1; PID:g2996

A/Experimental source: strain VFS

C/Genetics:

A/Gene: gcp

C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 21.0%; Score 290.5; DB 2; Length 335;  
 Best Local Similarity 34.3%; Pred. No. 1.8e-18;  
 Matches 85; Conservative 46; Mismatches 82; Indels 35; Gaps 10;

C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Jul-2004  
C:Accession: H64074  
R:Flitschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Galloway, J.D.; Scott, J.; Shirley, R.; Liu, L.L.; Glodde, A.; Kelley, J.M.; Weisman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrman, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter, A.  
Article: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: H64074  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-342 <TIG>  
A:Cross-references: UNIPROT:P43764; GB:U32735; GB:L42023; MID:g1573509; PIDN:AAC22187.1;  
C:Superfamily: O-sialoglycoprotein endopeptidase  
C:Keywords: hydrolase; metalloproteinase

Query Match	20.9%	Score 289.5	DB 2	Length 342
Best Local Similarity	35.7%	Pred. No. 2,3e-18		
Matches	82	Conservative	30	Mismatches 95; Indels 23; Gaps 7
QY	1	MEAHATRL-TNKVEPFLVLLISGGHLLAVOGSDFLILGSLDIAPGDMIDKAR	59	
DB	112	MEGHLLAPMLDNDSPHPFVALVSGGHTQVVRVGVKYEYTGESIDDAAGAADPKTK	171	
QY	60	RLSLIKPEGCSMGSKALIEHLAQGNRFHFDIKPLHAAKNCDFSFGTGHVTDKIWK	119	
DB	172	LLGL-DYF-----GGALSLRLAEKGTNRKTRTFPRPMIDRAGLDIFSGLKTFPAANTVQ	224	
QY	120	KEKEGG-IEKGQILLSSADIAATVQHTMACHLVKTRHAILFCQKRDILLPQNNALVNS	177	
DB	225	AKKEGELIEQ-----TKADIAVAFAQDVAVDTLA-----IKCK-RALKERYGRVLVA	271	
QY	178	GGVANSFPIRBALELTNAQTCTILCPBPRCTNGNMIANNGIERLBAQ	227	
DB	272	GGVANSKKRLRTLLAHLMONTLGGVAFYPPQPCQTDNGAMATAYGFLRLKQG	321	

RESULT 11  
AI0079  
probable glycoproteinase gcp [imported] - *Yersinia pestis* (strain CO92)  
C:Species: *Yersinia pestis*  
CjDate: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
A:Accession: AI0079  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchall, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; E  
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, F  
Native 413, 523-527, 2001  
A>Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586560  
A:Accession: AI0079  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC89500.1; PID:GL15978736; GSPDB:GN00175  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match Similarity 20.2%; Score 279.5; DB 2; length 337;  
Best Local Similarity 32.8%; Pred. No. 1.8e-17;  
Matches 81; Conservative 35; Mismatches 102; Indels 29; Gaps 7

```
Oy      1 MEAAATLTRL-TNKVEPFLVLLISGGHCLALVOGVSDPFLLGSLDAPGMDLVKAR 59  
        ||| ||||| :||| :||| :||| :||| :||| :||| :||| :  
Db     112 MEGHLADMLTEENADEPFVALVLVSGGHTOLIVTGICEYILLGSVDPAAGEAPDKTK 171  
  
Oy      60 RLSTLKHPBCSWMGSKAIEHLAQGNRFHPDIKPLHAANKCDSPFGOLGHVTQIKIM 119  
        :||| :||| :||| :||| :||| :||| :||| :||| :  
Db     172 LLGL-DYP-----GGPMLSRMAQQGYGRFTPPRPMPDRPGDLDFSGLKTPFAANTIRA 224  
  
Oy     120 KEKEGIEKGQILSSADIATATVQTMACHLVKRTHRALIFCKGDDLPÖNN-AVLVASG 178  
        ::||| :||| :||| :||| :||| :||| :||| :||| :
```

Db 225 NGDDD-----QTRADIARAFEDAVDTLAIKSKRA-----LDQTFKRLVIAAG 267  
Qy 179 GVASNFYIRALRLEITLNATQCTLLCPPLRLCTDNGIMIAMNGIERLRAGLGIHIDIEGIR 238  
Db 268 GVSANQTLRLKLAEMQKRGGEVFFYARPEFCTDNGAMIAVAGVRFKA--LNSLSVVS 324  
Qy 239 YEPKCP 245  
Db 325 VRPRWPL 331

## RESULT 12

AG0892  
probable glycoproteinase [imported] - Salmonella enterica subsp. enterica serovar Typhi (S  
C/Species: Salmonella enterica subsp. enterica serovar Typhi  
A/Note: this species has also been called Salmonella typhi  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C/Accession: AG0892  
R/Parikhili, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,  
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moulé, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A/Reference number: AB0502; MUID:21534947; PMID:11677608  
A/Accession: AG0892  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-337 <PAR>  
A/Cross-references: GB:AL513382; PIDN:CAD07733.1; PID:g16504285; GSPDB:GN00176  
C/Genetics:  
A/Gene: STY3387  
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 20.2%; Score 279.5; DB 2; Length 337;  
Best Local Similarity 32.9%; Pred. No. 1.8e-17;  
Matches 81; Conservative 35; Mismatches 103; Indels 27; Gaps 8;  
Qy 1 MEAHALITRL-TNKVEPFPVLISGGHCLALVGVSDPFLIGKSLDIAPGDMLEKVAR 59  
Db 112 MEGHLAPMLEBNPPFPFVALVSGHQLISVTGIGYELLGSDIDDAAGAPKTKAK 171  
Qy 60 RLSTLKHPBCSTWGGKALIEHLAKQGNRFPDIKPLHNAKNCDSFSTGLQHTDKITMK 119  
Db 172 LLGL-DYP-----GGPLLSKMAAOGTAGRFVPRPMTDRPGIDFSFGSKTFPAANTIRS 224  
Qy 120 KEKEGIEKGQILSSADIAATVOHTMACHLVKRTHRALIFCKQRDLPPONNAVIVASGG 179  
Db 225 NGDDE-----QTRADIARAFEDAVDTL-----MICK-RALSTGPKRLVMAAG 268  
Qy 180 VASNFYIRALRLEITLNATQCTLLCPPLRLCTDNGIMIAMNGIERLRAGLGIHIDIEGIRY 239  
Db 269 VSANRTLRALKLAEMQKRGGEVFFYARPEFCTDNGAMIAVAGVRFKA--GVTADL-GVTV 325  
Qy 240 EPKCP 245  
Db 326 RPRWPL 331

## RESULT 13

C91122  
probable O-sialoglycoprotein endopeptidase [imported] - Escherichia coli (strain O157:H7  
C/Species: Escherichia coli  
C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C/Accession: C91122  
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gaewara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A/Reference number: A53625; MUID:21156231; PMID:11258796  
A/Accession: C91122  
A/Status: preliminary  
A/Molecule type: DNA

A/Residues: 1-337 <HAY>  
A/Cross-references: GB:BA000007; PIDN:BA37370.1; PID:g13363420; GSPDB:GN00154  
A/Experimental source: strain O157:H7, substrain RIMD 0509952  
C/Genetics: EC83947  
A/Gene: EC83947  
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 20.1%; Score 278.5; DB 2; Length 337;  
Best Local Similarity 33.3%; Pred. No. 2.2e-17;  
Matches 82; Conservative 36; Mismatches 101; Indels 27; Gaps 9;  
Qy 1 MEAHALITRL-TNKVEPFPVLISGGHCLALVGVSDPFLIGKSLDIAPGDMLEKVAR 59  
Db 112 MEGHLAPMLEBNPPFPFVALVSGHQLISVTGIGYELLGSDIDDAAGAPKTKAK 171  
Qy 60 RLSTLKHPBCSTWGGKALIEHLAKQGNRFPDIKPLHNAKNCDSFSTGLQHTDKITMK 119  
Db 172 LLGL-DYP-----GGPLLSKMAAOGTAGRFVPRPMTDRPGIDFSFGSKTFPAANTY 222  
Qy 120 KEKEGIEKGQILSSADIAATVOHTMACHLVKRTHRALIFCKQRDLPPONNAVIVASGG 179  
Db 223 --RDNGTD-----QTRADIARAFEDAVDTLAIKSKRA-----DLTGPKR--LVWAGG 268  
Qy 180 VASNFYIRALRLEITLNATQCTLLCPPLRLCTDNGIMIAMNGIERLRAGLGIHIDIEGIRY 239  
Db 269 VSANRTLRALKLAEMQKRGGEVFFYARPEFCTDNGAMIAVAGVRFKA--GATADL-GVSV 325  
Qy 240 EPKCP 245  
Db 326 RPRWPL 331

## RESULT 14

B85967  
probable O-sialoglycoprotein endopeptidase ysjd [imported] - Escherichia coli (strain O1:  
C/Species: Escherichia coli  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C/Accession: B85967  
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
Miller, L.; Grobeck, B.J.; Davis, N.W.; Lin, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A/Reference number: AB5480; MUID:21074935; PMID:11206551  
A/Accession: B85967  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-337 <STO>  
A/Cross-references: GB:AB005174; NID:g12517643; PIDN:AG58198.1; GSPDB:GN00145; UMGD:2443  
A/Experimental source: strain O157:H7, substrain EDL933  
C/Genetics:  
A/Gene: ysjd  
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 20.1%; Score 278.5; DB 2; Length 337;  
Best Local Similarity 33.3%; Pred. No. 2.2e-17;  
Matches 82; Conservative 36; Mismatches 101; Indels 27; Gaps 9;  
Qy 1 MEAHALITRL-TNKVEPFPVLISGGHCLALVGVSDPFLIGKSLDIAPGDMLEKVAR 59  
Db 112 MEGHLAPMLEBNPPFPFVALVSGHQLISVTGIGYELLGSDIDDAAGAPKTKAK 171  
Qy 60 RLSTLKHPBCSTWGGKALIEHLAKQGNRFPDIKPLHNAKNCDSFSTGLQHTDKITMK 119  
Db 172 LLGL-DYP-----GGPLLSKMAAOGTAGRFVPRPMTDRPGIDFSFGSKTFPAANTY 222  
Qy 120 KEKEGIEKGQILSSADIAATVOHTMACHLVKRTHRALIFCKQRDLPPONNAVIVASGG 179  
Db 223 --RDNGTD-----QTRADIARAFEDAVDTLAIKSKRA-----DLTGPKR--LVWAGG 268  
Qy 180 VASNFYIRALRLEITLNATQCTLLCPPLRLCTDNGIMIAMNGIERLRAGLGIHIDIEGIRY 239  
Db 269 VSANRTLRALKLAEMQKRGGEVFFYARPEFCTDNGAMIAVAGVRFKA--GATADL-GVSV 325



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_blue\_p2n model

Run on: February 16, 2005, 13:19:07 ; Search time 3172.37 Seconds  
(without alignments)  
3203.653 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHALTRITNKVEPFLV.....DISKVEGASIKVQLKMEI 267

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=x1h  
-Q/cgm2.1/USPTO.spool/US10649273/runat.14022005.114703.16411/app.query.fasta.1.1429  
-DB=EST -QPM=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX=blowum62 -TRANS=humand40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPMT=pro -NOR=exc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10649273.QCGN.1.1.6799.@runat.14022005.114703.16411 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOBERRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1263	91.2	870	5	BQ423651 AGENCOURT
2	1224	88.4	2284	3	AK045669
3	1216	87.8	1622	3	AK011265 Mus muscu
4	1097	79.2	640	5	BQ636028 hd03d11.Y
5	1075	77.6	852	5	BX391919 BX391919
6	1071	77.3	1605	3	BC030671 Mus muscu
7	1053	76.0	701	2	BE740611 601595739
8	944	68.2	922	5	BQ961028 AGENCOURT
9	925	66.8	658	7	CF362328 829596 MA

10	916	66.1	822	7	CF257246
11	838	60.5	637	7	CK941819
12	830	59.9	597	6	CB272391
13	819.5	59.2	792	5	BU403563
14	816	58.9	730	7	CN823245
15	789	57.0	545	1	AV602901
16	786	56.8	866	5	BU127463 603114407
17	784	56.6	490	6	CB852881
18	749	54.1	736	7	CK365185
19	749	54.1	1173	6	CD508917
20	746	53.9	723	5	BU261251
21	722	52.1	812	5	BU246489
22	717	51.8	634	2	AM601179
23	709	51.2	484	1	AJ670918
24	708	51.1	736	6	CA057753
25	707.5	51.1	701	5	BU621780
26	704	50.8	749	7	CK982692
27	698	50.4	696	5	BU302606
28	692	50.0	579	7	CK819035
29	691.5	49.9	789	5	BU242187
30	683	49.3	909	5	BK756548
31	682	49.2	548	7	CO880741
32	671	48.4	682	2	BB043703
33	668	48.2	706	5	BU202465
34	666	48.1	869	5	BK754527
35	647	46.7	919	7	CF407294
36	646	46.6	878	5	BK776940
37	645	46.6	533	4	BM126453
38	638	46.1	1082	5	BK359023
39	625	45.1	506	2	BF415802
40	618.5	44.7	879	5	BU256052
41	614	44.3	861	5	BU246158
42	611	44.1	1171	5	BU261605
43	604	43.6	423	1	AA273921
44	602	43.5	414	1	AA589724
45	600	43.3	863	5	BU376295

## ALIGNMENTS

RESULT 1  
BQ423651  
LOCUS  
DEFINITION BQ423651 870 bp mRNA linear EST 23-MAY-2002  
AGENCOURT\_7790948 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6065828  
5', mRNA sequence.  
ACCESSION BQ423651 GI:21118966  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
EST.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 870)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DC/DTP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: llam13342 row: 1 column: 21  
High quality sequence stop: 710.  
Location/Qualifiers  
1..970  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6065828"

## FEATURES

source

/cissue type="melanotic melanoma"  
/lab host="DH10B (phage-resistant)"  
/clone lib="NIH MGC 72"  
/note="Organ: skin; Vector: pCMV-SPORT6, Site\_1: NotI,  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

## ORIGIN

## Alignment Scores:

Pred. No.:	1.9e-136	Length:	870
Score:	1263.00	Matches:	248
Percent Similarity:	98.42%	Conservative:	1
Best Local Similarity:	98.02%	Mismatches:	2
Query Match:	91.19%	Indels:	2
DB:	5	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x BQ423651 (1-870)

```

QY      1 MetGluAlaHisAlaLeuThrIleArgLeuPhrAsnValGluPhroPhleuVal 20
DB      |||
DB      112 ATGAGAGGCTGACGACTTACTATTAAGTTGACCAATTAAGTAATTTCTTTTATGTT 171
QY      21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
DB      |||
DB      172 CTTTGTGATTTCTGAGAGTCACTGCTGCTGTTGGCATTAAGTTCAAGAGCTTTCAGATTTCG 231
QY      41 LeuLeuGlyIleSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
DB      |||
DB      232 CTTCTTGAAAGTCTTTGACATGACCAAGGTGACATGCTTGAACAAGGTGGCAAGAAGA 291
QY      61 LeuSerLeuIleIleHisProGluCysSerThrMetSerGlyGlyValAlaIleGluHis 80
DB      |||
DB      292 CTTCTTTAAATAAATCAATGACAGAGTCTCCACATGAGTGGTGGGAAAGCCATTGAACAT 351
QY      81 LeuAlaIleGlnGlyAsnArgPheHisPheAspIleIleYsProProLeuHisHisAlaIle 100
DB      |||
DB      352 TTGGCCAAACAGCAAGAAATGATTTTATTTGATCAATCAACCTCCCTTGATATGCTTAA 411
QY      412 AATGTGATTTTCTTTACTGACACTTCAACGTTACTGATTAATATATATGAAAG 471
DB      |||
QY      121 GluIleGluGluGluGlyIleGluIleGlyGlnIleLeuSerSerAlaAlaAspIleAlaIa 140
DB      |||
DB      472 GAAAAAGAGAGAGATTTGAGAAAGGGGCAAAATCTGCTTCAGACAGACATTCGTGCC 531
QY      141 ThrValGlnHisThrMetAlaCysHisIleuValIleYsArgThrHisArgAlaIleLeuPhe 160
DB      |||
DB      532 ACAGTACAGCACACATATGCAATGCTTCTTGAAAAAGAACACATCGGCTATTTCTGTT 591
QY      161 CysIleGlnArgAspLeuLeuProGlnAsnAlaValIleuValAlaSerGlyVal 180
DB      |||
DB      592 TGTAGAGCAGAGACTTGTACTTCAAAATATATGACGATGCTGTTGATCTGCTGTC 651
QY      181 AlaSerAsnPheIleIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 200
DB      |||
DB      652 GCAAGTACTTCTATATCCGAGACCTCTGGAATTTTAAACAACGCAACACAGTGCAC 711
QY      201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleIleIleIleIle 220
DB      |||
DB      712 TTGGTGTGTCCTCTCCAGACTATGACCTGATTAATGATTAATGATTAATGATTAATGAT 771
QY      221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgIleGlu 240
DB      |||
DB      772 ATTGAAGACGACGCTGCTGGGCTTTTATCATGACATGAAGGACATCGGCTATGA 831
QY      241 -ProIleCysPro-LeuGlyValAspIleSerIle 251
DB      |||
DB      832 ACCAAATGTCCTCTTGAGATGACATATCAAAA 866

```

RESULT 2  
AK045669LOCUS  
DEFINITION

AK045669 2284 bp mRNA linear HTC 03-APR-2004  
Mus musculus adult male corpora quadrigemina cDNA, RIKEN  
full-length enriched library, clone: B30219017 Product: similar to  
PUTATIVE STADOLYLCOPOXYLASE TYPE 2 [Homo sapiens], full insert  
sequence.

ACCESSION  
VERSION  
KEYWORDS

AK045669.1 GI:26337528  
HTC; CAP trapper;  
Mus musculus (house mouse)

SOURCE  
ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636

REFERENCE  
AUTHORS

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
11042159

JOURNAL  
MEDLINE  
PUBMED

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsumoto, S.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multiplexed capillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

JOURNAL  
MEDLINE  
PUBMED

4 The RIKEN Genome Exploration Research Group Phase II Team and the  
PANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5

TITLE  
JOURNAL  
AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 2284)

JOURNAL  
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,  
Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kurata, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Subdivision

TITLE  
JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken

## COMMENT

Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>  
URL: <http://fantom.gsc.riken.jp/>  
Location/Qualifiers

1..2284

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="PANTOM\_DB:B230219017"

/db\_xref="taxon:10090"

/clone="B230219017"

/sex="male"

/tissue\_type="corpora quadrigemina"

/clone\_id="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

221..1465

/note="unnamed protein product; putative  
similar to PUTATIVE SIALOGLYCOPROTEIN TYPE 2 (Homo  
sapiens) (SPT2) (Q9H4B0, evidence: PASTY, 80%ID, 100%length,  
match=1242)"

/protein\_id="BAC32450.1"

/codon\_start=1

/db\_xref="GI:26337529"

/translation="MIMRRPRTGAIIPKPKRVGFLRSPVHPTLSCHLVGIER  
SCDTGAADVDEGTALGSAHSQTVHLKKGIVPFAQQLHRENTIRIVETLSAC  
RITPSDLISAIATTKIPGLALSIVGSLPSLDLVNQPKPFIPIHMEHALITRLTNK  
VEPPLVLLISGHCLLALVQGVSDFLIGSLDIPADMDIKVARSLIKHPEGST  
MSGKAIKDLADKGNRHFHTINPQWNAQDSFPTGQHLTDKLTKEKEGK  
OISSADIAAVOHAATACHLAKRTHAILCCKOKNLSPPANAVLVVSGVANSYIR  
KALEIVANATCTCLCPPRCTDNGIMIANNGIERLAAGLGVLDHVEDIRPKPL  
GVDSIRVAAEAIIVPLKML"

2262..2267

/note="putative"

2284

/note="putative"

polysignal  
polysite  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.75e-131 Length: 2284  
Score: 1224.00 Matches: 233  
Percent Similarity: 92.51% Conservative: 14  
Best Local Similarity: 87.27% Mismatches: 20  
Query Match: 88.38% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x AK045669 (1-2284)

1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal 20  
662 ATGAGGAGCTCAGCAGCTACTATTAGGCTCACCAATAAGTGAATTTCTTTTATGTT 721  
21 LeuLeuLeuSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
722 CTTTGTATTTCTGGCGGCTGCTGCTGTTGGCATTATGTCAGAGGTTGTCGATTTCTGTG 781  
41 LeuLeuGlyLysSerLeuAspIleAlaProGlyAspMetLeuAspLysValAlaArgArg 60  
782 CTCCTTGGAGAGCTTTTGGACATAGCACAGGCGACATGCTTGAACAGGTGCGAAGAGA 841  
61 LeuSerLeuIleLysAspProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHis 80  
842 CTTTCTTAAATCAACATCCAGAAATGTTCTCAATGAGAGTGGGAAAAGCTATAGACAG 901  
81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHisAlaLys 100  
902 TTGGCCAAAGACGAAATGATTCATTTTCTACTCAATCCCACTATGACAGATCTAAG 961  
101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys 120  
962 AATTGCAATTTTCTTCAACGGAGCTTCAACATATTACTGATGACTAATTAACACACAG 1021

QY 121 GlyLysGluGlnGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140  
DB 1022 GAAAAAGAAAGAGGACATTAGAAAGGCGCAATCTCTATCAGCTCAGACATGCTGCT 1081  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe 160  
DB 1082 GCGGTACAGACATGCAACAGCTGCCACTTGGCAAAAGAACACATCGCGCTATTCGTTT 1141  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 180  
DB 1142 TCCAGACAGAAAAAATTTGCTCTCCAGCTAACGCGATTAATTTATCTTGAAGGTGTT 1201  
QY 181 AlaSerAsnPheThrIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr 200  
DB 1202 GCAAGTAACTTGTATCATCCGAAAGCATTTGGAAATTTGCGAAATCCAGCGATGCACG 1261  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleAsnGly 220  
DB 1262 TTGTTGTGTCCACTCCCAAGACGTGCACATGACCAATGATCATGATTCATGATGATGGA 1321  
QY 221 IleGluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgGly 240  
DB 1322 ATGGAAGATTAAGTGTCTGCTGGCGCTGTTTACATGATGATGAGACATCCGATATGAA 1381  
QY 241 PolyLysCysProLeuGlyValAspIleSerLysGlnValGlyGluAlaSerIleLysVal 260  
DB 1382 CCAAAATGCTCTTGGAGATTAACATATCCAGAAAGATTGCGAAGCTGCATTAAGA 1441  
QY 261 ProGlnLeuLysMetGluIle 267  
DB 1442 CCGCATTAATAATGCACTT 1462

RESULT 3  
AK011265  
LOCUS 1622 bp mRNA linear HTC 03-APR-2004  
DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length  
enriched library, clone:2610001M19 product:similar to PUTATIVE  
SIALOGLYCOPROTEIN TYPE 2 (Homo sapiens), full insert sequence.  
ACCESSION AK011265  
VERSION AK011265.1 GI:12847275  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Komano, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
MEDLINE  
10349636  
PUBMED  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Komano, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
MEDLINE  
11042159  
PUBMED  
REFERENCES  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Komano, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Mochizuki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
MEDLINE  
11076861  
PUBMED

REFERENCE	4	The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
TITLE	5	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
JOURNAL		Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
REFERENCE	6	Nature 420, 563-573 (2002) (bases 1 to 1622)
AUTHORS		Adachi,J., Aizawa,K., Akahira,S., Altamura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Katsukawa,T., Kato,H., Kawai,U., Kojima,Y., Komou,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Shibata,K., Shidata,Y., Shinigawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Terima,Y., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE		Direct Submission
JOURNAL		Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT		Please visit our web site ( <a href="http://genome.gsc.riken.jp/">http://genome.gsc.riken.jp/</a> ) for further details.
		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAATTCACAGACTCTCTTTTCTTTTTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAATTCACAGATTAAATTAATATCCCCCCCCC 3']. cDNA was cleaved with XhoI and ScaI. Cloning sites, 5' end: XhoI; 3' end: ScaI. Host: SOUR.
FEATURES		location/Qualifiers
SOURCE	1..1622	/organism="Mus musculus"
		/mol_type="mRNA"
		/strain="C57BL/6J"
		/db_xref="PANTOM.DB:2610001M19"
		/db_xref="taxon:10090"
		/clone="2610001M19"
		/tissue_type="whole body"
		/clone_id="RIKEN full-length enriched mouse cDNA library"
		/dev_stage="10 days embryo"
	207..1451	
CDS		/note="unnamed protein product; putative similar to PUTATIVE SIALOGLYCOPROTEIN TYPE 2 [Homo sapiens] (SPTR Q9H4B0, evidence: FASTV, 80%ID, 100%length, match=1242)"
		/codon_start=1
		/protein_id="BAB27506.1"
		/db_xref="gi:12847272"
		/translation="MTMRRTKALIPPKSKYTGFLRRSVHPRTLSCHKLVIGITSCDTPGAADVDETNVNGEALHSQTVALTKGGIVPPVAQDLRENTORLVEETLSAKRIPSDSAIAITKPGIALSLISGVLFQLTVNQFPKPPTPIHHMAHALITRLTKRVSEFPAVLISGGCHLALVOGVDFPLDGLSIDLAGDMDDKHARLSILIKPECSMTWFKALIEOLAKGNRPHTNPMPNOKANDPSFTGOIHDTDLITHKEEGEIKEOISSADADIAAVGHATKCHARTKTHAILFECKKNLSPANNAVLYSGVASLTYYRKALISVANATCTCLCPRLCTONGIMIANWGIERLNAGVGVLHVEDIRYEKCPLVGDVSIREVAALIVPRLLKML"

polysignal	1605..1610	/note="putative"
polysite	1622	
polysite	/note="putative"	
Alignment Scores:	1.44e-130	Length: 1622
Pred. No.:	1216.00	Matches: 232
Score:	92.13%	Conservative: 14
Percent Similarity:	86.89%	Mismatches: 21
Best Local Similarity:	87.80%	Indels: 0
Query Match:	3	Gaps: 0
US-10-649-273-2_COPY_148_414 (1-267) x AK011265 (1-1622)		
QY	1 MetGluAlaHisAlaLeuThrIleArgLeuThrAsnLysValGluPheProPheLeuVal	20
DB	648 ATGAGAGCTACGACACGACATTAATAGCTCCACCAATAAAGAAATTTCCCTTTTAACTT	707
QY	21 LeuLeuIleSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu	40
DB	708 CTTTGAATTTCTGGCGGCTACCTGCTGTGGCATTTAGTCCAAAGGTGTTCCGATTTCTGT	767
QY	41 LeuLeuGlySerSerLeuAspIleAlaIleProGlySerPheLeuAspLysValAlaArgG	60
DB	768 CTCCTTGGGAAGTCTTTGGACATGACACACGACACATGCTTGAACAGGTGGCAAGAAG	827
QY	61 LeuSerLeuIleLysHisIleProGlyCysSerThrMetSerGlyGlyLysAlaIleGluHis	80
DB	828 CTTTCTTAACTCAACATCCGAAATGTTCTTCAATGAAGTGTGGAAAAGCTATAGAACAG	887
QY	81 LeuAlaLysGlnGlyAsnArgPheHisIlePheAspIleLysProProLeuHisIleAlaLys	100
DB	888 TTGGCCAAAGACGGAATAGATTCATTTTACTATCAATCCACCTATGACGAATGTAAAG	947
QY	101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysLys	120
DB	948 AATGCGCATTTTCTTTCACGGGACTTCMAATATTCATGATTAAGCTATATAACACACAA	1007
QY	121 GluLysGlnGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAla	140
DB	1008 GAAAAGAGAGAGGATTTGAGAGAGAGACAAATCTGTATCATAGCTGACGACATTCCTCT	1066
QY	141 ThrValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPhe	160
DB	1068 GCGGTACGACATGCAACAGCGGTGCACCTTGCGAAAAGAACACATCGCGTATTCGTCTT	1127
QY	161 CysLysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal	180
DB	1128 TGCAGACGAAAATTTGCTCTCTCCAGCTAACCAAGATTAAGTTGATTCGGAGGTGT	1187
QY	181 AlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThr	200
DB	1188 GCAAGTAATCTGTACATCCGAAAGACATTGGAAATTTGTCGCAATATGCAACGACGTGCAC	1247
QY	201 LeuLeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly	220
DB	1248 TTGTTGTGTCCACCTCCAGACGTGTGACCTACACATGACATCATGATTGCAATGAAATGA	1307
QY	221 IleGluArgLeuArgAlaGlyLeuGlyTyrlleLeuHisAspIleGluGlyTyrlleArgTyrlle	240
DB	1308 AATTAAAGATTAACGTGCTGCTGTGGCGCTTTTAACTGATGATGAAGAACATCCGATATAA	1367
QY	241 ProLysCysProLeuGlyValAspIleSerLysGlyValGlyGluAlaSerIleLysVal	260
DB	1368 CCAAAAGTCTCTTGTGAGACATATCCAGAAAGTTGACGAAGCTGCCATTAAGAATA	1427
QY	261 ProGlnLeuLysMetGluIle 267	
DB	1428 CCGCGATTAAATAATGCACTT 1448	

B0636028  
LOCUS B0636028 640 bp mRNA linear EST 15-JUL-2002  
DEFINITION h03d11.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he  
ACCESSION B0636028 Homo sapiens cDNA clone h03d11.5', mRNA sequence.  
VERSION B0636028.1 GI:21760487  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 640)  
Wistow, G., Bernslein, S.L., Wyatt, M.K., Ray, S., Behal, A.,  
Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K.,  
Expressed sequence tag analysis of human retina for the NEBank  
Project: Retbindin, an abundant, novel retinal cDNA and alternative  
splicing of other retina-preferred gene transcripts  
Mol. Vis. 8 (4), 196-204 (2002)  
JOURNAL MEDLINE  
PUBMED 12107411  
COMMENT Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 03 row: d column: 11  
Seq primer: M13R1 reverse primer (ABI).  
Location/Qualifiers  
1..640  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="h03d11"  
/tissue\_type="Retina"  
/dev\_stage="Adult"  
/lab\_host="EMDH10B"  
/clone\_lib="Human Retina cDNA (Un-normalized,  
unamplified): hd/he"  
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue  
was dissected from two 80 year old donors with no observed  
eye disease. 100ug of total RNA was used for library  
construction. A directionally cloned cDNA library in the  
pSPORT1 vector (Life Technologies) was constructed at  
Bioserve Biotechnology (Laurel MD) essentially following  
the protocols of the SuperScript Plasmid System full  
details of which are contained in the manufacturer's  
instruction manual (http://www.lifetech.com/). First  
strand synthesis was carried out using a Not I  
primer-adaptor  
[5'-pGACTAGTTCAGATCCGACGCGCCG(T)15-3']. EST analysis  
was performed on the unamplified library at the NIH  
Intramural Sequencing Center (NISC)."

ORIGIN  
Alignment Scores:  
Pred. NO.: 2.94e-117 Length: 640  
Score: 1097.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 79.21% Indels: 0  
DB: 5 Gaps: 0  
US-10-649-273-2\_COPY\_148\_414 (1-267) x B0636028 (1-640)

QY 59 AArgatgleuSerleuIlelyshisProgluCyserThmetSerGlylyalaalle 78  
Db 3 AGAAGACTTCTTTATTAATCAATCATCCAGTCTCCACATCATAGTGTGGAAAGCCATA 62  
QY 79 GtuhstleuAlaIyGlnGlyasnArgPheHispeAspIlelyeProProleuHis 98  
Db 63 GAACATTGTGGCCAAACAGAAATAGATTTCATTGTGACATCAAACTCCCTTGCAATCAT 122

QY 99 AAlaYasnCyseAspPheSerPheThrGlyleuGlnHisValThrAspLyalleIleMet 118  
Db 123 GCTAAATAATTTGATATTTCTTTTACTGACCTCAACACGTTACTGATAAATATATATG 182  
QY 119 lYelYsgIuYsgIuGlyGlyIleGlyYsgIyGlnIleleuSerSerAlaIaaspIle 138  
Db 183 AAAAAGCAAAAGAGAGAGATTTGAGAGGGGCAATCTGCTTCAGACAGACATTT 242  
QY 139 AAlaIatThValGlnHisThrMetAlaCyseHisleuVallyArgThrHisArgAlaIle 158  
Db 243 GCTGCCACAGTACAGACACAAATGGATGATCATCTGTGAAAGAACACATCGGCTATTT 302  
QY 159 leuPheCyseYsgIuArgAspIleuProgluAsnAsnAlaValleuValaIaSerGly 178  
Db 303 CTTGTTTGAAGAGAGAGACTTTGTTACTCAAAATATACAGTACTGTTGATCTGCT 352  
QY 179 GtYValAlaSerAsnPheTyrlleaArgAlaIeGlyIleleuThrAsnAlaThrGln 198  
Db 363 GGTGTGCAAGTAACTTATATATCCGACAGCTCTGGAATTTTAACAACGACACAG 422  
QY 199 CyThrleuLeuCyseProProProlArgleuCyThrAspAsnGlyIleMetIlealTrp 218  
Db 423 TGCACCTTGTGTGTCTCTCTCCACAGCTATGACATGATATGCAATTATGATTGATG 482  
QY 219 AangIyIleGlyArgleuArgAlaGlyIleuGlyIleleuHisAspIleGlyIleArg 238  
Db 483 AATGGATTGAAGAAAGACTACGCTGCTGGGCTTTTACATATCATATGAAGGCATCCG 542  
QY 239 TYRGluprolyeCyseProleuGlyValaAspIleSerlyGlyValGlyIuAlaSerIle 258  
Db 543 TATGAACCAAAATGTCCTCTTGAGATGACATATCAAAAGACTTGGAGAGCTTCATA 602  
QY 259 lYsValProGlnleuIyMetGluIle 267  
Db 603 AAAGTACCAAAATTAATAATGAGATA 629

RESULT 5  
BX391919/c 852 bp mRNA linear EST 28-APR-2004  
LOCUS BX391919 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
DEFINITION cDNA clone CS0DK001Y802 3-PRIME, mRNA sequence.  
ACCESSION BX391919  
VERSION BX391919.2 GI:46846154  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 852)  
Li, W.B., Gruber, C., Jessee, J. and Polyes, D.  
Full-length cDNA libraries and normalization.  
Unpublished (2001)  
On May 13, 2003 this sequence version replaced gi:30611736.  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: segre@genoscope.cns.fr, web: www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
1240.r  
For more information about this cluster, see  
http://www.genoscope.cns.fr/cdna?c=CS0BA1035E07\_CS03317\_1&c=1240.r

FEATURES  
source Location/Qualifiers  
1..852  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DK001Y802"



/cell\_type="HELA CELLS COT 25-NORMALIZED"  
/cell\_line="HELA"  
/clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

## Alignment Scores:

Pred. No.:	1,65e-114	Length:	852
Score:	1075.00	Matches:	208
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	77.62%	Indels:	0
DB:	5	Gaps:	0

US-10-649-273-2\_COPY\_148\_414 (1-267) x BX391919 (1-852)

QY	1	MetGuaIAHsAlaLeuThrlleArGleuThraNlyVaIGlupheProheLeuV	20
Db	626	ATGAGGCTCATGCACTTACTTATAGTTGACCAATTAAGTGAATTTCTTTTATTT	567
QY	21	LeuLeuIleSerGIyGIyHISCySleuLeuAlaLeuValGInGIyValSerAspPheLeu	40
Db	566	CTTTTGAATTTCTGAGAGTCATGTCGTTGGCATTAGTTCAAGAGTTTCAGATTTTCG	507
QY	41	LeuLeuGIyLySerLeuAspIleAlaPArgIyAspMetLeuAspLyValAlaArg	60
Db	506	CTTCTTGAAAGTCTTTTGACATACACACAGGTGACCTTGCACAGGTGCAGAAAGA	447
QY	61	LeuSerLeuIleLyHISProGIyCySerThrMetSerGIyGIyLyValIleGIyHIS	80
Db	446	CTTTCTTTAAATTAACATTCACAGTCTCACCATGAGTGGGAAAGCCATAGAAACT	387
QY	81	LeuAlaLyGIyGIyAsnArPheHISpHeAspIleLySProProLeuHISHisAlaLyS	100
Db	386	TTGGCCAAACAGAAATATGATTTTCATTTTGACATCAACCTCCCTTGATCATGCTAA	327
QY	101	AsnCyAspPheSerPheThrGIyLeuGIyHISValThrAspLySleIleMetLyS	120
Db	326	AATGTGTATTTTCTTTTACTGACCTTCAACGCTTACTGATTAATATATGAAG	267
QY	121	GIyLySGIyGIyGIyIleGIyLyGIyHISLeuSerSerAlaAlaAspIleAla	140
Db	266	GAAAGAGAGAGATTAAGAGAGGCAATCTGCTTCAAGCAGCAGCAATGCTGCC	207
QY	141	ThrValGIyHISThrMetAlaCySHISLeuValLySArgThrHISArgAlaIleLeuPhe	160
Db	206	ACAGTACGACACACATGTCATCTTGTGAAAGAACACATCGGCTATTTCTGTT	147
QY	161	CysLySGIyAspLeuLeuProGIyHISAsnAlaValLeuValAlaSerGIyGIyVal	180
Db	146	TGTATGACGAGAGACTTGTACTCAAAATATATGACATGCTGTCATCTGTGTGTC	87
QY	181	AlaSerAsnPheTyrlleArgAlaLeuGIyIleLeuThrAsnAlaThrGIySThr	200
Db	86	GCAAGTAACTTCTATATCCGAGAGCTCTGAAATTTTAAACAAACGACACAGTGCAC	27
QY	201	LeuLeuCySProProProArgLeu	208
Db	26	TTGTGTCTCTCTCTCCACGACTA	3

RESULT 6  
LOCUS BC030671 1609 bp mRNA linear HTC 19-NOV-2003  
DEFINITION Mus musculus O-6-ialoglycoprotein endopeptidase-like 1, mRNA (cDNA  
clone IMAGE:1226118), containing frame-shift errors.  
ACCESSION BC030671  
VERSION BC030671.1 GI:21040459  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1609)  
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shmumen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.F., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, Y., Hsieh, P.,  
Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,  
Schetz, T.E., Brownstein, M.J., Uedini, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Kaha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huylk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahney, J., Helton, B., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitt, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smallus, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
2 (bases 1 to 1609)  
Straussberg, R.  
Direct Submission  
Submitted (20-May-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Marcello Bento Soares, Ph.D.  
CDNA Library Preparation: M. Bento Soares, University of Iowa  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadan@systemsbiology.org](mailto:amadan@systemsbiology.org)  
Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting

REMARK  
COMMENT

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>  
Series: IRAX Plate: 66 Row: e Column: 10  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 2112463  
This clone has the following problem: frame shifted.  
Location/Qualifiers  
1..1609  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1226118"  
/tissue\_type="Thymus gland, mouse"  
/clone\_lib="Soares\_thymus\_2NMNT"  
/lab\_host="DH10B"  
/note="Vector: pTT73-Pac"

FEATURES  
source

## ORIGIN

## Alignment Scores:

Pred. No.:	1,21e-113	Length:	1609
Score:	1071.00	Matches:	211
Percent Similarity:	83.52%	Conservative:	12
Best Local Similarity:	79.03%	Mismatches:	20
Query Match:	77.33%	Indels:	24
DB:	3	Gaps:	1

US-10-649-273-2\_COPY\_148\_414 (1-267) x BC030671 (1-1609)

```

Qy      1 MetGlua1aHis1a1aLeuThr11leaRgLeuThrAenLysValGluPheProPheLeuVal 20
Db      665 ATGAGAGGCTCAGCAGTCACTATTAGGCTCACCACATTAAGTGAATTTCTTTTAGTT 724
Qy      21 LeuLeu11eserGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      725 CTTTGGATTTCCTGGCGGCTCAGCTGCTGTGGCATATGTCACAGAGTGTTCCGATTTCCTG 784
Qy      41 LeuLeuGlyLysSerLeuAsp11lea1ProGlyAspMetLeuAspLysValAlaArgArg 60
Db      785 CTCTTGGAGAACTCTTGGACATAGCACAGCGCAGCATGCTTGACAAAGGTGGCAAGAAGA 844
Qy      61 LeuSerLeu11elyHis1aProGlyCysSerThrMetSerGlyGlyValAla11leGluHis 80
Db      845 CTTTCTTTAATCAAAATCCAGAAATGTTCTACATAGAGTGGGAAAGCTTATAGACAG 904
Qy      81 LeuAlaLysGlnGlyAsnArgPheHisPheAsp11elysProProLeuHis1a1aLys 100
Db      905 TTGGCCAAAGACGGAATAATGATTCATTTACTATCAATCCACTATGACAAATGCTAAG 964
Qy      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLys11le11eMetLysLys 120
Db      965 AATTGCCAATTTTCTTTCACCGGACCTTCAACATATTAAGTATTAACATTAACACACAG 1024
Qy      121 GluLysGlnGluGly11leGluLysGlyGln11leLeuSerSerAlaAlaAsp11lea1a1a 140
Db      1025 GAAAGAAAGAAAGGCAATTTGAGAAAGGAGCAATCTGTCTACCTGCTGACAGCATTTCTGCT 1084
Qy      141 ThrValGlnHis1a1aLeuThr11leaRgLeuValLysArgThrHisArgAla11leLeuPhe 160
Db      1085 GCGGTACAGCATGCAACAGCGTGCACCTTGGAAAGAACACATCGCGCTATTCTGTTT 1144
Qy      161 CysLysGlnArgAspLeuLeuProGlnAsnAlaVal11leValAlaSerGlyGlyVal 180
Db      1145 TGCACAGCAAGAAATTTGCTCTCTCCAGCTAACGCAATTAATGTTATCTGAGAGTGT 1204
Qy      181 AlaSerAspPheThr11leaRgArgAlaLeuGlu11leLeuThrAsnAla11leThrGlnCysThr 200
Db      1205 GCAAGTAATCTGTATCATCTCCGAAGCATTTGGAATTTGCGCAATATGCAACGAGTGCACG 1264
Qy      201 LeuLeuCysProProArgLeuCysThrAspAsnGly11leMet11a1a11leArgAsnGly 220
Db      1265 TTGTTGTCTCCACTCCAGACTGTGACTGACATGCAATGGCATATATTTGCA----- 1315
Qy      221 11leGluArgLeuArgAlaGlyLeuGly11leLeuHisAsp11leGluGly11leArgTrGlu 240
Db      1315 ----- 1315
Qy      241 ProLysCysProLeuGlyValAsp11leSerLysGlyValGlyLysVal11leLysVal 260
Db      1316 ---TGAATGTCCTCTTGGAGTACATATCCAGAGAGTTGCAAGAGTGCCTAATAAGTA 1372
Qy      261 ProGlnLeuLysMetGlu11le 267
Db      1373 CCGCGATTAAATAATGCACTT 1393

```

```

RESULT 7
BE740611 701 bp mRNA linear EST 15-SRP-2000
LOCUS BE740611 601595739P1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3949640 5',
DEFINITION mRNA sequence.
ACCESSION BE740611.1 GI:10154603
VERSION BE740611.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 701)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

```

```

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: DCTD/DP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: L1CM814 row: n column: 09
High quality sequence stop: 701.
Location/Qualifiers
1..701
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3949640"
/issue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 9"
/note="Organ: ovary; Vector: pORF7; Site 1: XhoI; Site 2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCAAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

## ALIGNMENT SCORES:

```

Pred. No.: 4,66-112 Length: 701
Score: 1053.00 Matches: 206
Percent Similarity: 99.05% Conservative: 3
Best Local Similarity: 97.63% Mismatches: 0
Query Match: 76.03% Indels: 2
DB: 2 Gaps: 0

```

US-10-649-273-2\_COPY\_148\_414 (1-267) x BE740611 (1-701)

```

Qy      1 MetGlua1aHis1a1aLeuThr11leaRgLeuThrAenLysValGluPheProPheLeuVal 20
Db      71 ATGAGAGGCTCAGCAGTCACTATTAGGCTCACCACATTAAGTGAATTTCTTTTAGTT 130
Qy      21 LeuLeu11eserGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40
Db      131 CTTTGGATTTCCTGGCGGCTCAGCTGCTGTGGCATATGTCACAGAGTGTTCCGATTTCCTG 190
Qy      41 LeuLeuGlyLysSerLeuAsp11lea1ProGlyAspMetLeuAspLysValAlaArgArg 60
Db      191 CTCTTGGAGAACTCTTGGACATAGCACAGGATGACATCTTGAACAGAGGCGCAAGAGA 250
Qy      61 LeuSerLeu11elyHis1aProGlyCysSerThrMetSerGlyGlyValAla11leGluHis 80
Db      251 CTTTCTTTAATCAAAATCCAGAAATGTTCTACATAGAGTGGGAAAGCTTATAGACAT 310
Qy      81 LeuAlaLysGlnGlyAsnArgPheHisPheAsp11elysProProLeuHis1a1aLys 100
Db      311 TTGGCCAAAGACGGAATAATGATTCATTTGACATCAACCTCCCTTGATCATGCTAAA 370
Qy      101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLys11le11eMetLysLys 120
Db      371 AATTGCCAATTTTCTTTCACCGGACCTTCAACATATTAAGTATTAACATTAACACACAG 430
Qy      120 GluLysGlnGluGly11leGluLysGlyGln11leLeuSerSerAlaAlaAsp11lea1a1a 140
Db      431 GGAACAAAGAAAGAGTATTTGAGAGGCGCAATCTGTCTTACAGCAGACATTTGCTG 490
Qy      140 1aThrValGlnHis1a1aLeuThr11leaRgLeuValLysArgThrHisArgAla11leLeuP 160
Db      491 CCACAGTACAGACACATATGCAATCTCTTGTGAAAAGACACATCGCGGATATCTGT 550
Qy      160 heCysLysGlnArgAspLeuLeuProGlnAsnAlaVal11leValAlaSerGlyGlyVal 180

```

Db 551 TTGTAAAGCAGAGACTGTGTAACCTCAAAATATATGACAGTACGTGGTGCATCTGGTGG 610  
Qy 180 AAlaSerAspPheTyrTlleArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysT 200  
Db 611 TCGCAAGTAACTCTTAATATCCGACAGAGCTCTGGAATTTTAAACAACGACACAGTGA 670  
Qy 200 hrLeuLeuCySPProProArgLeuCyS 209  
Db 671 CTTTGTGTGTCTCTCTCCACAGCTATGCG 699  
RESULT 8  
LOCUS B0961028 922 bp mRNA linear EST 21-AUG-2002  
DEFINITION AGNCOURT 8863711 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:6423902  
5' mRNA sequence.  
ACCESSION B0961028.1 GI:22376506  
VERSION B0961028.1 GI:22376506  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 922)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: DCTD/DRP/Gazdar  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LHC2603 row: d column: 15  
High quality sequence stop: 584.  
Location/Qualifiers  
1..922  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6423902"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_18"  
/note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Library constructed by King Hong in the  
Laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

ORIGIN  
Alignment Scores:  
Pred. NO.: 3.65e-99 Length: 922  
Score: 944.00 Matches: 207  
Percent Similarity: 89.45% Conservative: 5  
Best Local Similarity: 87.34% Mismatches: 16  
Query Match: 68.16% Indels: 10  
DB: 5 Gaps: 3  
US-10-649-273-2\_COPY\_148\_414 (1-267) x B0961028 (1-922)

Qy 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnIleValGlnPheProhLeuVal 20  
Db 207 ATGAGGCTCATGCACTAATTAAGGTGACCAATAAGTAATTCCTTTTATGTT 266  
Qy 21 LeuLeuIleSerGlnGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
Db 267 CTTTGAATTTCTGAGAGTCACTGCTGTGGCAATTAAGTTCAAGAGCTTCAAGATTTTCTG 326

Qy 41 LeuLeuGlnIlySerSerLeuAspIleAlaProGlnYAspMetLeuAspIlyValAlaArgArg 60  
Db 327 CTTCTTGGAAAGCTCTTTGACATAGACACAGGATGACATCTTGACAAAGGTGGCAAGAA 386  
Qy 61 LeuSerLeuIleIlyHisProGlnCys-SerThrMetSerGlnGlyIlyAlaIleGlnIle 80  
Db 387 CTTTCTTTAATAAACATCCAGATGACATCCATGATGATGGGGAAGCCATAGACA 446  
Qy 80 bleuAlaIlyGlnGlnIlyAsnArgPheHisPheAspIleIlyProProLeuHisAlaIly 100  
Db 447 TTTGGCCAAACAGGAATAGATTTCATTTTACATCAACCTCCCTTCATCATCTTA 506  
Qy 100 sAsnCyAspPheSerPheThrGlnIlyLeuGlnHisIlyValThrAspIlyIleIleMetIly 120  
Db 507 AATATTGTATTTTCTCTTACTGACCTTCACACCTTACTGATTAATATATATGAAAA 566  
Qy 120 sGlnIlySGlnGlnGlnIlyIleGlnIlyGlnIlyLeuSerSerAlaAlaAspIleAla 140  
Db 567 GGAAGAAAGAGAGATTTGAGAAAGGGAATCCGTCTTCAGACGACAGACTTGTCTGC 626  
Qy 140 aThrValGlnHisThrMetAlaCysHisLeuValIlyAsnArgThrHisArgAlaIleLeuPh 160  
Db 627 CACAGTACAGCACACAAATGCAATGCAATCTTGTGAAAAAGAAA-CATCGGGCTAATCTGTT 685  
Qy 160 eCysIlySGlnArgAspLeuProGlnAsnAsnAlaValLeuValAlaSer-GlnGlyIly 180  
Db 686 TTGTAAAGCAGAGACTTGTACTCTCAAAATATATGACAGTACGTGGTGCATCTTGGGGGG 745  
Qy 180 AAlaSerAspPheTyrTlleArgAlaLeuGlnIleLeuThr-AsnAlaThrGlnCys 199  
Db 746 TCGCAAGTAACTCTTAATATCCGCAACCTCTGGAATAATTTACAAACCCACACGGGGC 805  
Qy 200 Thr-LeuLeuCySPProProArgLeuCySPThr-AspAsnGlnIlyMetIleAlaTTP- 218  
Db 806 ACCTTGTGTGGGTCCCGCCCAACTATGACTGTGATATGGG---CATTTATGATGCG 862  
Qy 219 -----AsnGlnIlyIleGlnArg---LeuArgAlaGlnIlyLeuGlnIly 229  
Db 863 TGGGGAAGGGAATTGAAAAAATAACCTGCTGCGCTTGGGG 903  
RESULT 9  
LOCUS CF362328/6 658 bp mRNA linear EST 25-AUG-2003  
DEFINITION CF362328 MARC 3P1G Sus scrofa cDNA 3', mRNA sequence.  
ACCESSION CF362328  
VERSION CF362328.1 GI:34161882  
KEYWORDS EST.  
SOURCE Sus scrofa (pig)  
ORGANISM Sus scrofa  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
REFERENCE 1 (bases 1 to 658)  
AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Valliet,J.L., Fox,J.,  
Wise,T.A., Noneman,D.J., Wray,J.B. and Keefe,J.W.  
TITLE A second set of porcine ESTs from a pooled-tissue normalized  
library  
JOURNAL Unpublished (2003)  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called with phred v0.020425.c and  
trimmed with the aid of the trim\_alt option. Vector identified with  
cross\_match v0.990329.  
Plate: SR8015 row: H column: 12  
Seq primer: TAGAAGGACAGTCGAGG.  
Location/Qualifiers  
1..658  
/organism="Sus scrofa"  
/mol\_type="mRNA"

/db\_xref="taxon:9823"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 3P1G"  
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;  
Library made with RNA pooled from multiple tissues  
including brain, liver, muscle, placenta/endometrium,  
ovary, testes, and bone marrow."

## ORIGIN

Alignment Scores:  
Pred. No.: 3,71e-97 Length: 658  
Score: 925.00 Matches: 174  
Percent Similarity: 93.97% Conservative: 13  
Best Local Similarity: 87.44% Mismatches: 12  
Query Match: 66.79% Indels: 0  
DB: 7 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x CF362328 (1-658)

QY 20 ValLeuLeuLeuSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPhe 39  
DB 657 GTTCTTTGATATCTGGCGTCATTGCTTTGGCATTAGAGAGAGTTTCAGATTCTT 598  
QY 40 LeuLeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArg 59  
DB 597 CTGCTTCTTGACAGCTCTTTGGACATACACAGGTGACATGCTTGACAGGTGACAGCA 538  
QY 60 ArgLeuSerLeuIleLeuHisProGlyCysSerThrMetSerGlyGlyValAlaIleGlu 79  
DB 537 AGACTTCTTAAATAAACATCCAGAGTGTCCACATGAGTGGTGGAGGAGCCATTAGAA 478  
QY 80 HisLeuAlaLeuGlnGlyValAsnArgPheHisPheAspIleLeuProLeuHisIleAla 99  
DB 477 CATTGTCCAAACAGGGAATAGTGAATTTGATTGCAACCTCCATCCATGCAAGTAT 418  
QY 100 LysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetLys 119  
DB 417 AAAATATGATGATTTCTTTCTGACATCAACAGTATTTGATTAAGACAAATTAAGCAG 358  
QY 120 LysGlnLysGlnGlnGlyIleGlnLysGlnIleLeuSerSerAlaAlaAspIleAla 139  
DB 357 AAGGAAAAAGGAGAAAGTATGAGAAAGGCGCAATCTGCTTCAGCTGACAGACTTGT 298  
QY 140 AlaThrValGlnHisThrMetAlaCysHisLeuValIleValArgThrHisArgAlaIleLeu 159  
DB 297 GCTCAGATGACGACACAGTACCTGCGCATTTGCAAAAAGAACTCATGTCATTCGTG 238  
QY 160 PheCysLysGlnArgAspLeuLeuProGlnAsnAlaValIleLeuValAlaSerGlyGly 179  
DB 237 TTTTGCAAAACGAGAGACTTATATGATGCAAGTAAATGAGATTTGTTGATCTGAGGT 178  
QY 180 ValAlaSerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCys 199  
DB 177 GTTGCAAGTAACTTATATATGAGAAAGCTTTAGAAAGTGTGACAAAGTCAACAAATCC 118  
QY 200 ThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrp 218  
DB 117 ACTGCTGTGTCTCTCTCCAGACTATGACATGATTAATGCGATTATGATTCATGCG 61

## RESULT 10

CF257246

CF257246 822 bp mRNA linear EST 07-AUG-2003  
LOCUS ph008 g02 PHA-activated splenocytes Gallus gallus cDNA, mRNA

DEFINITION

ACCESSION  
CF257246  
VERSION  
CF257246.1 GI:33490501

KEYWORDS

SOURCE  
EST.  
Gallus gallus (chicken)ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 822)  
AUTHORS Wittzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whitaker,C.J.,  
Chausse,A.M. and Zoorob,R.  
TITLE A collection of chicken ESTs from activated immune cells  
JOURNAL Unpublished (2003)  
COMMENT Contact: Zoorob R  
UPR 1983  
CNRS

7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France

Tel: 33 1 49 58 35 00  
Fax: 33 1 49 58 33 81  
Email: zoorob@vjf.cnrs.fr.

Location/Qualifiers

FEATURES  
source 1..822  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/cell\_type="Splenocytes"  
/clone\_lib="PHA-activated splenocytes"  
/note="Vector: pTriplex2"

## ORIGIN

Alignment Scores:  
Pred. No.: 5,75e-96 Length: 822  
Score: 916.00 Matches: 175  
Percent Similarity: 81.37% Conservative: 39  
Best Local Similarity: 66.54% Mismatches: 49  
Query Match: 66.14% Indels: 1  
DB: 7 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x CF257246 (1-822)

QY 1 MetGlnAlaHisAlaLeuThrIleArgLeuThrAsnLysValGlnPheProPheLeuVal 20  
DB 35 ATGGAGGCGCCACAGCACTTATCCATGACATGACAGACAGAAATTAATTCCTTCTAGTT 94  
QY 21 LeuLeuLeuSerGlyGlyHisCysLeuLeuAlaLeuValGlnGlyValSerAspPheLeu 40  
DB 95 CTTTACTCTCCGAGAGTACTGCACTTGGCAGTGCAGAGAGAGTTTCAGATTCTCTT 154  
QY 41 LeuLeuGlyLeuSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60  
DB 155 CTGCTTGACACATCCATGATATACAGACACAGGTGACATTTGATTAAGTATGACAAAG 214  
QY 61 LeuSerLeuIleLeuHisProGlyCysSerThrMetSerGlyGlyValAlaIleGluHis 80  
DB 215 CTCTCTTATGTAAGACACCCGAGATGCCACATGATGCCGCGGAGGAAAGCAATAGAGCAC 274  
QY 81 LeuAlaLysGlnGlyAsnArgPheHisPheAspIleLeuProLeuHisIleAlaLys 100  
DB 275 CTGGCTCAAAACCGAGACTGGCAACAGTACCTTTCAGCTTCCATCCATGCAAGTATCGT 334  
QY 101 AsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspIleIleMetLysLys 120  
DB 335 AACTGTATTTTCTTTCTCCGACTTCAGACCTTGTCAACAAAGCAATCTTCAGAA 394  
QY 121 GlnLysGlnGlnGlyIleGlnLysGlnIleLeuSerSerAlaAlaAspIleAla 140  
DB 395 GAAAAAGAAAGAGTATTCAGAAAGGGAATCTGTCTCGCGTTAAGACATCGCTGCT 454  
QY 141 ThrValGlnHisThrMetAlaCysHisLeuValIleValArgThrHisArgAlaIleLeuPhe 160  
DB 455 GCTGCAACGACAGTGTGCTGCTCTATATATTCAGACGACACACCGAGCCATGCTCTTC 514  
QY 161 CysLysGlnArgAspLeuLeuProGlnAsnAlaValIleLeuValAlaSerGlyGlyVal 180  
DB 515 TGCATGAAAACAGCATATTTATCCAAAACCTGCACTCTGTGTATCAGAGAGATT 574  
QY 181 AlasGlnHisThrTyrlleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysTrp 200  
DB 575 GCAATTAATCACTATATATGAGAAAGAGCTGCAAGACTCTGCAAAATGCAAAAGTTTGTCT 634  
QY 201 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 220

Db 635 TTCTGCTCTCTCCAGAGCTGTGACCGATTAATGATGATTCAGATGAC 694  
Qy 221 TTTCTGCTCTCTCCAGAGCTGTGACCGATTAATGATGATTCAGATGAC 240  
Db 695 ATTAAAGGTTGCGTGCAGATGATGATTTTATTCAGTACGATCGCTACGAA 754  
Qy 241 ProlyscProLeuGlyValAspIleSerLySgluValGlyGluAlaSerIleVal 260  
Db 755 CCAAAAGCTCCCTTGGATGATGATTTTCCAAAGGTTGAA-GAGGATTCATCAAGT 813  
Qy 261 ProGlnLeu 263  
Db 814 CCAAGACTA 822  
RESULT 11  
CK941819/c 637 bp mRNA linear EST 15-MAR-2004  
LOCUS 4065407 BARC 10BOV Bos taurus cDNA clone 10BOV12\_F24 3', mRNA  
DEFINITION  
ACCESSION CK941819 GI:45456199  
VERSION CK941819  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 637)  
AUTHORS Sonstegard, T.S., Van Tassel, C.P., Matukumalli, L.K., Harhay,  
G.P., Bock, S., Rubenfield, M. and Gasparre, L.C.  
TITLE Production of EST from cDNA libraries derived from immunologically  
activated bovine gut  
JOURNAL Unpublished (2004)  
COMMENT Contact: Tad S. Sonstegard  
Bovine Functional Genomics Laboratory  
Animal and Natural Resources Institute  
Bldg. 200 RMA BARC-East, Beltsville, MD 20705, USA  
Tel: 3015048416  
Fax: 3015048414  
Email: tads@nri.barc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred  
0.000925 using options -trim alt -c rim fasta. Vector identified  
by cross match using options -mismatch 12 -minscore 18  
Plate: 12 row: F column: 24  
Seq primer: AGCGATACAAATTCACACAGG  
High quality sequence stop: 637.  
Location/Qualifiers  
1. 637  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/strain="Holstein"  
/db\_xref="taxon:9913"  
/clone="10BOV12\_F24"  
/sex="Male"  
/issue\_type="Pooled"  
/dev\_stage="Multiple"  
/lab\_host="DH10B T1 phage resistant"  
/clone\_id="BARC 10BOV"  
/note="Organ: Small Intestine; Vector: pagen-1; Site 1:  
ECOFV; Organ: NotI; Equimolar amounts of mRNA extracted  
from proximal jejunums of 18 and 21 wk old steers, and  
distal ileums of 14 day old calves. Proximal jejunum  
exposed to C. oncophora for 3 and 6 weeks, and distal  
ileum exposed to C. parvum for 7 days"

## ORIGIN

Alignment Scores:  
Pred. No.: 5.09e-87 Length: 637  
Score: 838.00 Matches: 153  
Percent Similarity: 92.47% Conservative: 19  
Best Local Similarity: 82.26% Mismatches: 14  
Query Match: 60.51% Indels: 0

DB: 7 Gaps: 0  
US-10-649-273-2\_copy\_148\_414 (1-267) x CK941819 (1-637)  
Qy 82 AATySGInGlyAsnArgPheHisPheAspIleLysProProLeuHisSalAlaLysAsn 101  
Db 632 GCCAAACAGGGAAATAGATTCATTTGATTTCCAGCTCCCAATGACAGCTGCTAAAAAT 573  
Qy 102 CysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysGlu 121  
Db 572 TGTGATTTTCTTTTCTGAGCTCAACAGTTATTTGATTAAGATGATTAATGAAAAAGAA 513  
Qy 122 LysGluGluGlyIleGluLysGlyGlnIleLysSerSerAlaAlaAspIleAlaIleThr 141  
Db 512 AAAGAGAAAGTATTCAGACAGGCGAGCTCTCTCTTCAGCTGACAGCATGCTGTGGC 453  
Qy 142 ValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCys 161  
Db 452 GTGACAGCACACCGTGGCTGCGACATTCGCAAAAGAACACATCGCTCTCTGTCTGCG 393  
Qy 162 LysGlnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAla 181  
Db 392 AAGCAGAGAGGCTTCTTACGTACAGATACGAGTACGTGTGTATCTGTAGAGCGCTGCA 333  
Qy 182 SerAsnPheTyrlleArgArgAlaLeuGlnIleLeuThrAsnIleThrGlnCysThrLeu 201  
Db 332 AGTAACTTATATATTCGAAAGACCTTGAAATTTGACCAATGCAACACAGTGCATTTG 273  
Qy 202 LeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIlePheAsnGlyIle 221  
Db 272 CTGTGCGCGCCCCCAGACTTTCAGCTGACAGAGCGGCTATGATTCAGTAAGATGTGTT 213  
Qy 222 GluArgLeuArgAlaGlyLeuGlnIleLeuHisAspIleGluGlyIleArgTyrgluPro 241  
Db 212 GAAAGACTACCGTGTGCTGTGGCATTTTACCAACACAGAGGACATCCGCTACAGAACCA 153  
Qy 242 LysCysProLeuGlyValAspIleSerLySgluValGlyGluAlaSerIleLysValPro 261  
Db 152 AATGTCTCTTGGATGATATATCAAAAGATTTGGAGAACTCTATTAAGTCCCA 93  
Qy 262 GlnLeuLysMetGluIle 267  
Db 92 AGATTAAAAATGAGATT 75  
RESULT 12  
CB272391 597 bp mRNA linear EST 24-FEB-2003  
LOCUS ma157912.y1 McCarrey Eddy spermatoocytes Mus musculus cDNA clone  
DEFINITION IMAGE:645750 5' similar to TR:Q9VWD6 Q9VWD6 CG14231 PROTEIN. ;,  
mRNA sequence.  
ACCESSION CB272391 GI:28462714  
VERSION CB272391.1  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 597)  
AUTHORS McCarrey, J., Eddy, M., Marra, M., Hillier, L., Clifton, S., Pape, D.,  
Martin, J., Wyllie, T., Dante, M., Bowers, Y., Theising, B., Gibbons, M.,  
Ritter, E., Tsagarelis, R., Ronko, I., Maguire, L., Kennedy, S.,  
Bennett, J., Waterston, R. and Wilson, R.  
TITLE NIHES Mouse  
JOURNAL Unpublished (2002)  
COMMENT Contact: McCarrey/Eddy NIHES Mouse  
NIHES Mouse  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library constructed and donated by J. McCarrey, Ph.D. (Southwest  
Foundation for Biomedical Research, Dept. of Genetics) - excision

done by E.M. Eddy, Ph.D. (National Institutes of Health, National  
Institute of Environmental Health Sciences).

MG1:2069710

Seq primer: Primer name ambiguous

High quality sequence stop: 419.

## FEATURES

## SOURCE

```

1. 597
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:6445750"
/sex="male"
/tissue_type="spermatocytes, pooled from multiple mice"
/dev_stage="60 day"
/lab_host="DH10B (phage-resistant)"
/clone_lib="McCarrey Eddy spermatocytes"
/notes="Organ: testis; Vector: pBluescript SK+
(Stratagene); Site 1: XhoI; Site 2: EcoRI; cDNA oligo
dt-primed (5'-(GA)10-ACGAGCTCGAGTTTCTTTT-3') and
directionally cloned using 5' linkers 5'-AATCGCAGAG-3'
and 5'-CTCGTCCG-3'. Size selection of >400bp material
gives average insert size ranging from 1-2 kb. Library was
mass excised (from lambda-UniZAP-XR) and resulting
single-stranded phagemids were prepped and retransformed
into DH10B. Library contains 96% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63422."

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 3.98e-86 Length: 597
Score: 830.00 Matches: 162
Percent Similarity: 91.80% Conservative: 6
Best Local Similarity: 88.52% Mismatches: 15
Query Match: 59.93% Indels: 0
DB: Gaps: 0

```

US-10-649-273-2\_COPY\_148\_414 (1-267) x CB272391 (1-597)

```

QY 1 MetGUAhahisAlaleuthrIleatrgleuthrAnilyValglupheProPhleuVal 20
DB 49 ATGAGAGCTCAGCAGCTGATATTAGGCTCACCAATTAAGTAGAATTTCTTTTATGTT 108
QY 21 leuLeuileSerGlyGlyYhiScyLeuLeuAlaleuValGinglyValSerAspHeu 40
DB 109 CTTTGAATTTCTGGCGGTCTGCTGCTGTCATTAAGTCAAGAGGTGTTCCGATTTCTGTG 168
QY 41 leuLeuGlyYssSerLeuAspIleAlaProGlyAspMetLeuAspIleValAlaArgArg 60
DB 169 CTCCTTGGAGAGTTTGTGACATATGACACAGGCAATGCTTGACAAAGGTGCAAGAA 228
QY 61 leuSerleuileYhiAspProGlyCysSerThrmSerGlyGlyValAlaIleGluHis 80
DB 229 CTTTCTTAATCAAAATCCAGATGTTCTTACATGATGCTGGAAGAGCTATGAGAAG 288
QY 81 leuAlaIysGlnGlyAsnArgPheIlePheAspIleYsProProLeuHihisAlaLys 100
DB 289 TTGGCCCAAGACGGAATATGATTCATTTTACTATCAATCCATATGCAAGATGCTAAG 348
QY 101 AsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspYsIleIleMetLysLys 120
DB 349 AATTGCAATTTTCTTTCACGGGACTTCAACATATTAAGCTAATATACACACAG 408
QY 121 GlnYsGlnGlyGlyIleGlnLysGlyGlnIleLeuSerSerAlaAlaAspIleAla 140
DB 409 GAAAAAAGAGAGCATTTGAGAGAGGGCAATCTGTATCATGCTGACAGCATTCCTCT 468

```

```

QY 141 ThrValGlnHisThrmMetAlaCysHisLeuValYsArgThrHisArgAlaIleLeuPhe 160
DB 469 GCGGTACACATGCCACACACAGCTGCCACTTGGCAAAAGAACACATGCGGTATTCGTTT 528
QY 161 CysYsGlnArgAspLeuLeuProGlnAsnAlaValleuValAlaSerGlyGlyVal 180
DB 529 TCCAAGCAGAAAAAATTTCTCTCTCCAGCTAACGCAAGTATTAAGTTATCTGAGGTGTT 588
QY 181 AlaSerAsn 183
DB 589 GCAAGTAAC 597

```

## RESULT 13

## BU043563

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

1 (bases 1 to 792)  
Boardman, P.E., Sanz-Bzquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
A Comprehensive Collection of Chicken CDNAS  
Curr. Biol. 12 (22), 1965-1969 (2002)  
2235534  
12445392  
Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers

## FEATURES

## source

```

1. 792
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, HiseX"
/db_xref="taxon:9031"
/clone="CHST967014"
/dev_stage="F36"
/lab_host="DH10B"
/clone_lib="CSBQCHN59"
/notes="Organ: limbs; Vector: pBluescript II KS(+); Site 1:  
EcoRI; Site 2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was bluntended, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

```

## ORIGIN

## Alignment Scores:

```

Pred. No.: 1.01e-84 Length: 792
Score: 313.50 Matches: 164
Percent Similarity: 78.99% Conservative: 39
Best Local Similarity: 63.81% Mismatches: 53
Query Match: 59.17% Indels: 3
DB: Gaps: 1

```



ORGANISM

Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.

REFERENCE

1 (bases 1 to 545)  
Takauega,A., Hirotsune,S., Itoh,R., Jitchono,A., Suzuki,H., Aso,H.  
and Sugimoto,Y.  
Establishment of a high throughput EST sequencing system using  
poly(A) tail-removed cDNA libraries and determination of 36,000

TITLE

Bovine ESTs

JOURNAL

Nucleic Acids Res. 29 (22), E108 (2001)

MEDLINE

21570554

PUBMED

11713328

COMMENT

Contact: Yoshikazu Sugimoto  
Animal Genetics Division  
Shirakawa Institute of Animal Genetics  
Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan  
Tel: 81-248-25-5641  
Fax: 81-248-25-5725  
Email: kazusugi@cocoa.ocn.ne.jp  
Single pass sequencing.  
This clone was obtained from a polyA-deleted cDNA library.  
Location/Qualifiers

FEATURES

source

1..545  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/clone="E1K1013A07"  
/tissue\_type="kidney"  
/dev\_stage="fetus"  
/lab\_host="DH108"  
/clone\_lib="Bos taurus kidney fetus"  
/note="Vector: pZL1; Site\_1: SalI; Site\_2: NotI; Poly A  
was deleted from a NotI site"

ORIGIN

Alignment Scores:

Pred. No.: 2.14e-81 Length: 545  
Score: 789.00 Matches: 147  
Percent Similarity: 90.11% Conservative: 17  
Best Local Similarity: 80.77% Mismatches: 17  
Query Match: 56.97% Indels: 1  
DB: 1 Gaps: 0

US-10-649-273-2\_COPY\_148\_414 (1-267) x AV602901 (1-545)

QY 85 G|YAsnArgPheHisPheAspIlelySProPLeuHisHis-AtAltyAsnCyAspPh 104  
DB 544 GGAATAGATTCATTTGATTTCCANCTCCATGCAACGGTCTAAAAATTGTGATTT 485  
QY 104 eSerPheThrGlyLeuGlnHisValThrAspLyseIleIleMetLySlyGluLyGluG1 124  
DB 484 TTTCTTTCTGCACTTCAACAGTTATTGATTAAGATGATTAATGCAAAAGAAAANAGGA 425  
QY 124 uGlyIleGlyLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHi 144  
DB 424 AGGTATGAGAGAGGGGCGGTCTGTCTTCAGCTCGGACATTCGCTTGCGCGTCCAGCA 365  
QY 144 sThMetAlCyHisIleuValLySArgThrHisArgAlaIleLeuPheCySlyGlnAr 164  
DB 364 CACGGTGCCCTGGCAGATTCGAAAGAACACATGCTCTTCTGTTCTGCAAGCAGAG 305  
QY 164 gAspLeuLeuProGlnAsnAsnAlaValIleuValAlaSerGlyIValAlaSerAsnPh 184  
DB 304 AGGCTTCTTACATCAGAGTAACGAGTACTGTTGTATCTGAGGCGTCGCAAGTAACTT 245  
QY 184 eTyTlIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCySthrLeuLeuCySpr 204  
DB 244 ATATATCCGAAAGCCCTGGAAATGTGTACCAATGCAACAGTGCACATTGTGTGTCG 185  
QY 204 oProProArgLeuCyThrAspAsnGlyIleMetIleAlaTTPaAsnGlyIleGluArgLe 224  
DB 184 GCCCCCACTGACTGCACTGCAACAGCGGTATGATTCATGAGATGATGTTGAAAGACT 125

QY 224 uArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyGluProLySlySpr 244  
DB 124 AGTGCTGAGCTTGCGCATTTTACACACACAGAGGCGATCCGTACGAAACCAAAATGTC 65  
QY 244 oLeuGlyValAspIleSerLySgluValGlyGluAlaSerIleLyValProGlnLeuLy 264  
DB 64 TCTTGAGTATGATATATATCAAAAGAAAGTGGAGAGCTGCTATAAAAGTCCCAAGATTAAA 5  
QY 264 sMet 265  
DB 4 AATG 1

Search completed: February 16, 2005, 21:04:47  
Jod time : 3184.37 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 16, 2005, 10:46:17 ; Search time 45.2739 Seconds  
(without alignments)  
3019.954 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_148\_414  
Perfect score: 1385  
Sequence: 1 MEAHALTRILNKVEPPFLV.....DISKVEGRASIKVPOLKMEI 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1385	100.0	414	2	Q96EV9 homo sapien
2	1358.5	98.1	439	2	Q9H4B0
3	1231	88.9	414	2	Q6PEB4
4	1224	88.4	414	2	Q6BLB6
5	1216	87.8	414	2	Q9D0N0
6	1116	80.6	467	2	Q6AYN7
7	1105	79.8	364	2	Q96NH5
8	859	62.0	404	2	Q8JFW3
9	855	61.7	404	2	Q8JFR7
10	453.5	32.7	401	2	Q7Q9I8
11	415.5	30.0	323	2	Q960S6
12	415.5	30.0	409	2	Q9VMD6
13	407.5	29.4	480	2	Q22145
14	369	26.6	335	2	Q73H71
15	356.5	25.7	360	2	Q92LH8
16	352	25.4	387	2	Q98E16
17	349	25.2	387	2	Q68XR3
18	342	24.7	251	2	Q93FJ2
19	339.5	24.5	389	2	Q6ND54
20	338	24.4	365	2	Q8UC47
21	338	24.4	366	2	Q7CWB8
22	338	24.4	387	1	GCP_RICPR
23	334.5	24.2	359	2	Q8YJ11
24	334.5	24.2	359	2	Q8YJ11
25	333.5	24.1	344	2	Q8YJ11
26	329.5	23.8	344	2	Q8YJ11
27	326.5	23.6	357	2	Q8YJ11
28	326.5	23.6	357	2	Q8YJ11
29	321.5	23.2	346	2	Q6GIR3
30	320	23.1	364	2	Q6GIR3
31	317	22.9	340	2	Q6FCX9

32	316.5	22.9	346	2	Q7W668
33	315.5	22.8	346	2	Q7W134
34	304.5	22.0	343	2	Q9CJ11
35	304	21.9	353	2	Q7WQ09
36	302.5	21.8	341	2	Q7WDE3
37	302	21.8	339	2	Q6LV10
38	299.5	21.6	355	2	Q677H2
39	298	21.5	364	2	Q6EYF1
40	293	21.2	344	2	Q65RP0
41	291.5	21.0	341	2	Q915V7
42	290.5	21.0	335	1	GCP_AQUAE
43	289.5	20.9	342	1	GCP_HABTN
44	286.5	20.7	337	2	Q82XN2
45	285.5	20.6	348	2	Q917A5

## ALIGNMENTS

RESULT 1	PRELIMINARY	PRT	414 AA.
Q96EV9			
Q96EV9			
01-DEC-2001 (TREMBLrel. 19, Created)			
01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
OSGEPPL protein.			
Homo sapiens (Human).			
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
NCBI_Taxid=9606;			
[1]			
SEQUENCE FROM N.A.			
TISSUE=Muscle;			
MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnae.242603899;			
Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,			
Klausner R.D., Collins F.S., Wagner L., Shenn C.M., Schler G.D.,			
Altschul S.F., Zeeberg B., Burow K.H., Schefer C.F., Bhat N.K.,			
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
Stapleton M., Soares M.B., Bonaldo M.F., Casarini P., Prange C.,			
Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Scheetz T.E.,			
Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
Boak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,			
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,			
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,			
Krzywinski W.I., Skalska U., Small D.E., Scherch A., Schein J.E.,			
Jones S.J., Marra M.A.			
"Generation and initial analysis of more than 15,000 full-length human			
and mouse cDNA sequences."			
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
[2]			
SEQUENCE FROM N.A.			
TISSUE=Muscle;			
Strausberg R.L.			
Submitted (JUL-2001) to the EMBL/GenBank/DBS databases.			
EMBL: BC011904; AAH11904.1; --			
MEROPS: M22.004; --			
GO: GO:0008450; F-O-sialoglycoprotein endopeptidase activity; IEA.			
GO: GO:0008270; F-zinc ion binding; IEA.			
GO: GO:0006508; P-proteolysis and peptidolysis; IEA.			
InterPro: IPR009055; Peptidase M22.			
InterPro: IPR009180; Pept M22_Osialgl.			
Pfam: PF00814; Peptidase M22; 1.			
PIRSP: PIRSP04537; Osialgl; 1.			
PRINTS: PR00789; OSIALOPTASE.			
ProDom: PD002367; Peptidase M22; 1.			
TIGRfam: TIGR00329; gcp: 1_			
SEQUENCE 414 AA; 45122 MW; A536B333F5C68BD CRC64;			

Query Match 100.0%; Score 1385; DB 2; Length 414;

Best Local Similarity 100.0%; Pred. No. 2.9e-110; Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MEAHATLIRLTNRYEPFVLLISGHCILALVOGVSDFLLGKSLDIAFGMDLKVARR 60  
 DB 148 MEAHATLIRLTNRYEPFVLLISGHCILALVOGVSDFLLGKSLDIAFGMDLKVARR 207  
 QY 61 LSLIKHPCSTMSGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYTDKIMKK 120  
 DB 208 LSLIKHPCSTMSGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYTDKIMKK 267  
 QY 121 EKEGIEKQILSSADIAATVQHTMACHLVKTRAILFCQKORLLPQNNNAVVASGCV 180  
 DB 268 EKEGIEKQILSSADIAATVQHTMACHLVKTRAILFCQKORLLPQNNNAVVASGCV 327  
 QY 181 ASNFYIRRAEILTNATQCTLLCPPEPLCTDNGIMIANNGLERLHAGILHIDIGIRYE 240  
 DB 328 ASNFYIRRAEILTNATQCTLLCPPEPLCTDNGIMIANNGLERLHAGILHIDIGIRYE 387  
 QY 241 PKCPLGVDISKEVGASIKVPOLKMEI 267  
 DB 388 PKCPLGVDISKEVGASIKVPOLKMEI 414

## RESULT 2

Q9H4B0 PRELIMINARY; PRT; 439 AA.

AC Q9H4B0; 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Putative sialoglycoprotein type 2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Chen J.M., Fortunato M., Barrett A.J.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AJ295148; CAC14666.1; --  
 DR MEROPS; M22.004; --  
 DR Genew; HNC:23075; OSGEPL1.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR00905; Peptidase M22.  
 DR InterPro; IPR009180; Pept M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptide; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR Prodom; PD002367; Peptidase\_M22; 1.  
 DR TrRFams; TIGR00329; gcp; 1.  
 KW Protease.

SEQUENCE 439 AA; 48040 MW; 448493720784841F CRC64;

Query Match 98.1%; Score 1358.5; DB 2; Length 439;

Best Local Similarity 91.1%; Pred. No. 5.8e-108; Matches 266; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 MEAHATLIRLTNRYEPFVLLISGHCILALVOGVSDFLLGKSLDIAFGMDLKVARR 60  
 DB 148 MEAHATLIRLTNRYEPFVLLISGHCILALVOGVSDFLLGKSLDIAFGMDLKVARR 207  
 QY 61 LSLIKHPCSTMSGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYTDKIMKK 120  
 DB 208 LSLIKHPCSTMSGKALIEHLAKQGNRPHFDIKPPLHAKNCDPSFTGLQHTYTDKIMKK 267  
 QY 121 EKEGIEKQILSSADIAATVQHTMACHLVKTRAILFCQKORLLPQNNNAVVASGCV 180

DB 268 EKEGIEKQILSSADIAATVQHTMACHLVKTRAILFCQKORLLPQNNNAVVASGCV 327

QY 156 RAILFCQKORLLPQNNNAVVASGCVASNFYIRRAEILTNATQCTLLCPPEPLCTDNGIM 215

DB 328 RAILFCQKORLLPQNNNAVVASGCVASNFYIRRAEILTNATQCTLLCPPEPLCTDNGIM 387

QY 216 IANNGLERLHAGILHIDIGIRYEPKCPPLGVDISKEVGASIKVPOLKMEI 267

DB 388 IANNGLERLHAGILHIDIGIRYEPKCPPLGVDISKEVGASIKVPOLKMEI 439

## RESULT 3

Q6P6B4 PRELIMINARY; PRT; 414 AA.

AC Q6P6B4; 05-JUL-2004 (TREMBLrel. 27, Created)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaller G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Miliaty S.J.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska J., Smallos D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP SEQUENCE FROM N.A.  
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;  
 RA Strausberg R.;  
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC058172; AAH58172.1; --  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR00905; Peptidase M22.  
 DR InterPro; IPR009180; Pept M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptide; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR Prodom; PD002367; Peptidase\_M22; 1.  
 DR TrRFams; TIGR00329; gcp; 1.  
 KW Hypothetical protein.

SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBACB CRC64;

Query Match 88.9%; Score 1231; DB 2; Length 414;

Best Local Similarity 87.3%; Pred. No. 4.6e-97; Matches 233; Conservative 15; Mismatches 19; Indels 0; Gaps 0;

QY 1 MEAHATLIRLTNRYEPFVLLISGHCILALVOGVSDFLLGKSLDIAFGMDLKVARR 60

Db 148 MEAHLATRLTNKVEPFPVLVLLISGHCLLAVOGVSDPFLLGKSLDIPAGMDLVKVAR 207  
 Qy 61 LSLIKHEPCSTMSGKAI EHLAKOGNRFHDIKPLHAKNCDPFTGLOHTDITKIMX 120  
 Db 208 LSLIKHEPCSTMSGKAI EHLAKOGNRFHDIKPLHAKNCDPFTGLOHTDITKIMX 267  
 Qy 121 EKEBIEKQIILSSADIAATVOTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 180  
 Db 268 EKEBIEKQIILSSADIAATVOTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 327  
 Qy 181 ASNFYIRALLETNAOTCTLLCPPLCTDNGIMIANNGIERLPAAGILHDIGIRYE 240  
 Db 328 ASNFYIRALLETNAOTCTLLCPPLCTDNGIMIANNGIERLPAAGILHDIGIRYE 387  
 Qy 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267  
 Db 388 PKCPLGVDISKEVGEASIKVPOLKMEI 414  
 RESULT 4  
 Q8BLB6 PRELIMINARY; PRT; 414 AA.  
 AC Q8BLB6; PRELIMINARY; PRT; 414 AA.  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone: B230219017 product: similar to PUTATIVE STALOOLYCOPROTEASE TYPE 2.  
 DE Name=Osgsep11, GN  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/350555500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Komno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa S., Nagaoka S., Sasaki N., Carninci P., Komno H., Akiyama J., Nishi K., Kitamura T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto K., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Okazaki Y., Ishikawa T., Ozawa K., Tanaka T., Matsuzura S., Kawai J.,  
 RA "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RA Aichi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hirotsu T., Hirozane T.,  
 RA Hori F., Imotani K., Ishii Y., Ito M., Kagawa I., Katsukawa T.,  
 RA Kaich H., Kawai J., Kojima Y., Kondo S., Komno H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai K., Sakai C., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toyota T., Yasuniishi A., Muramatsu M., Hayashizaki Y.,  
 RL Submitted (Jul-2001) to the EMBL/GenBank/DBD databases.  
 DR EMBL; AK045669; BAC32450.1; --  
 DR MEROPS; M22.004; --  
 DR MGD; MGI:1919335; Osgsep11.  
 DR GO; GO:0008450; P:O-Glycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0008508; F:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009095; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialgl\_ptlde; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 KW Peptidase.  
 SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;  
 Query Match 88.4%; Score 1224; DB 2; Length 414;  
 Best Local Similarity 87.3%; Pred. No. 1.8e-96;  
 Matches 233; Conservative 14; Mismatches 20; Indels 0; Gaps 0;  
 Qy 1 MEAHLATRLTNKVEPFPVLVLLISGHCLLAVOGVSDPFLLGKSLDIPAGMDLVKVAR 60  
 Db 148 MEAHLATRLTNKVEPFPVLVLLISGHCLLAVOGVSDPFLLGKSLDIPAGMDLVKVAR 207  
 Qy 61 LSLIKHEPCSTMSGKAI EHLAKOGNRFHDIKPLHAKNCDPFTGLOHTDITKIMX 120  
 Db 208 LSLIKHEPCSTMSGKAI EHLAKOGNRFHDIKPLHAKNCDPFTGLOHTDITKIMX 267  
 Qy 121 EKEBIEKQIILSSADIAATVOTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 180  
 Db 268 EKEBIEKQIILSSADIAATVOTMACHLVKTRTRAILFCQKRDLLPQNNAVLVASGV 327  
 Qy 181 ASNFYIRALLETNAOTCTLLCPPLCTDNGIMIANNGIERLPAAGILHDIGIRYE 240  
 Db 328 ASNFYIRALLETNAOTCTLLCPPLCTDNGIMIANNGIERLPAAGILHDIGIRYE 387  
 Qy 241 PKCPLGVDISKEVGEASIKVPOLKMEI 267  
 Db 388 PKCPLGVDISKEVGEASIKVPOLKMEI 414  
 RESULT 5  
 Q8DON0 PRELIMINARY; PRT; 414 AA.  
 AC Q8DON0; PRELIMINARY; PRT; 414 AA.  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone: 2610001M19 product: similar to PUTATIVE STALOOLYCOPROTEASE TYPE 2.  
 DE

GN Name=Osgep11;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Mech. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20108560; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA The FANTOM Consortium,  
 RT the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayashizaki Y., Sugahara Y., Shibata K., Itoh M.,  
 RA Komano H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtractions of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Komano H., Akiyama J., Nishi K., Kikunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama S., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,  
 RA Okazaki Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multichannel sequencing.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Hanagaki T., Hara A., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito R., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamoto T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AK011265; BAB27506.1;  
 DR MEROPS: M22.004;  
 DR MGD: MG11919335; Osgep11.  
 DR GO: GO:0008450; F-O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO: GO:0008233; F:peptidase activity; IEA.  
 DR GO: GO:0008270; F:zinc ion binding; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR000905; Peptidase\_M22.

DR InterPro: IPR009180; Pept M22\_Osialg1.  
 DR Pfam: PF00814; Peptidase\_M22; 1.  
 DR PIRSF: PIRSF004537; Osialg1\_ptide; 1.  
 DR PRINTS: PR00789; OS1MOPYASE.  
 DR ProDom: PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs: TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SEQUENCE 414 AA; 44999 MW; 999BC689944DDB24 CRC64;  
 Query Match 87.8%; Score 1216; DB 2; Length 414;  
 Best Local Similarity 86.9%; Pred. No. 8.9e-96;  
 Matches 232; Conservative 14; Mismatches 21; Indels 0; Gaps 0;  
 QY 1 MEAHLATRLTKVPEPFVLLISGGHCLTAVGVSDPLLGKSLDIPAGMDLVAR 60  
 DB 148 MEAHLATRLTKVPEPFVLLISGGHCLTAVGVSDPLLGKSLDIPAGMDLVAR 207  
 QY 61 LSLIKHPGCSWGGKALEHLAKQGNRFHDIKPLPHAKNCDPSFTGLQHYTDKIMRK 120  
 DB 208 LSLIKHPGCSWGGKALEHLAKQGNRFHDIKPLPHAKNCDPSFTGLQHYTDKIMRK 267  
 QY 121 EKEBGEIKQIISADIAATVQHTMACHLVTRTRAILFCQRIIDLPNNAVLYASGV 180  
 DB 268 EKEBGEIKQIISADIAATVQHTMACHLVTRTRAILFCQRIIDLPNNAVLYASGV 327  
 QY 161 ASNFYIRALETLTATQCTLLCPPRLCTDNGIMIANGIERLRAGGLIDISIRYE 240  
 DB 328 ASNFYIRALETLTATQCTLLCPPRLCTDNGIMIANGIERLRAGGLIDISIRYE 387  
 QY 241 PKCPPLGVDSKEVGEASIKVPOLKMEI 267  
 DB 368 PKCPPLGVDSKEVGEASIKVPOLKMEI 414  
 RESULT 6  
 O6AYN7 PRELIMINARY; PRT; 467 AA.  
 AC O6AYN7;  
 DT 25-OCT-2004 (TREMBLrel. 28, Created)  
 DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.C.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
 RA Diatchenko L., Murnusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley D.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield J.S.,  
 RA Krzywinski M.I., Skalska U., Smalios D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;

RA Director MGC Project;  
 RL Submitted (Aug-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC078974; AAT8974.1; -  
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO: GO:0008270; F:zinc ion binding; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR00905; Peptidase M22.  
 DR InterPro: IPR009180; Pept M22 Osa1gl.  
 DR InterPro: IPR002016; Peroxidase.  
 DR Pfam: PF00814; Peptidase\_M22; 1.  
 DR PIRSF: PIRSF004537; Osa1glc\_pptds; 1.  
 DR PRINTS: PR00789; OSIALOPTASE.  
 DR ProDom: PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs: TIGR00329; gcp; 1.  
 DR PROSITE: PS00435; PEROXIDASE\_1; UNKNOWN\_1.  
 KM Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 467 AA; 50799 MW; 474E18B1959B8AC0 CRC64;  
 Query Match 80.6%; Score 1116; DB 2; Length 467;  
 Best Local Similarity 88.4%; Pred. No. 3.8e-87;  
 Matches 214; Conservative 10; Mismatches 18; Indels 0; Gaps 0;  
 Db 1 MEAHALTIRLTNKVPPPLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLVARR 60  
 148 MEAHALTIRLTNKVPPPLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLVARR 207  
 QY 61 LSLIKHPECSMGSGKAI EHLAKQGNRPHFDIKPPLHAKNCDPFTGLQHTVDTKIIMKK 120  
 Db 208 LSLIKHPECSMGSGKAI EHLAKQGNRPHFTINPMQNAKNCDFPTGLQHTVDTKIIMKK 267  
 QY 121 EKEEGIEKQILSSADIAATVQHTMACLVKTRTRAILFCQKORDLIPONNAVIVASGV 180  
 Db 268 EKEEGIEKQILSSADIAATVQHTMACLVKTRTRAILFCQKXVLSPPNAVIVASGV 327  
 QY 181 ASNFYIRRALIITLVNATQCTLLCPPRRLCTDNGIMIANNGIERLAGLIDIEGIRYE 240  
 Db 328 ASNFYIRRALIITLVNATQCTLLCPPRRLCTDNGIMIANNGIERLAGLIDIEGIRYE 387  
 QY 241 PK 242  
 Db 388 PK 389  
 RESULT 7  
 Q96NH5 PRELIMINARY; PRT; 364 AA.  
 AC Q96NH5;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein FLJ30879.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,  
 Murakami K., Yasuda T., Iwayanagi T., Magatsuma M., Shiraori A.,  
 Muraaki K., Hoshino T., Kaku Y., Kodaïra H., Kondo H., Sugawara M.,  
 Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
 Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki K.,  
 Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
 Tanai H., Kimura M., Watanabe M., Hiroaka S., Chiba Y., Ishida S.,  
 Oono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
 Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,  
 Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

RA Musashino K., Yunki H., Oshima A., Sasaki N., Aotsuka S.,  
 RA Yoshikawa Y., Matsumura H., Ichihara T., Shiohata N., Sano S.,  
 RA Moriya S., Momiyama H., Saich N., Takami S., Terashima Y., Suuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujisawa T.,  
 RA Oono T., Yamada K., Fujii Y., Ozaki K., Hirao Y., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togsashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Sato H., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 DR EMBL: AK055441; BAB70923.1; -  
 DR MEROPS: M22.004; -  
 DR GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO: GO:0008270; F:peptidase activity; IEA.  
 DR GO: GO:0008233; F:peptidase activity; IEA.  
 DR GO: GO:0008270; F:zinc ion binding; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR00905; Peptidase M22.  
 DR InterPro: IPR009180; Pept M22 Osa1gl.  
 DR Pfam: PF00814; Peptidase\_M22; 1.  
 DR PIRSF: PIRSF004537; Osa1glc\_pptds; 1.  
 DR PRINTS: PR00789; OSIALOPTASE.  
 DR ProDom: PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs: TIGR00329; gcp; 1.  
 KM PROSTATE.  
 SQ SEQUENCE 364 AA; 39528 MW; E0E605A07D0ECAD6 CRC64;  
 Query Match 79.8%; Score 1105; DB 2; Length 364;  
 Best Local Similarity 98.6%; Pred. No. 2.5e-86;  
 Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Db 1 MEAHALTIRLTNKVPPPLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLVARR 60  
 148 MEAHALTIRLTNKVPPPLVLLISGGHCLALVQGVSDPFLLGKSLDIAPGMDLVARR 207  
 QY 61 LSLIKHPECSMGSGKAI EHLAKQGNRPHFDIKPPLHAKNCDPFTGLQHTVDTKIIMKK 120  
 Db 208 LPLIKHPECSMGSGKAI EHLAKQGNRPHFDIKPPLHAKNCDPFTGLQHTVDTKIIMKK 267  
 QY 121 EKEEGIEKQILSSADIAATVQHTMACLVKTRTRAILFCQKORDLIPONNAVIVASGV 180  
 Db 268 EKEEGIEKQILSSADIAATVQHTMACLVKTRTRAILFCQKORDLIPONNAVIVASGV 327  
 QY 181 ASNFYIRRALIITLVNATQCTLLCPPRRLCTDNGIMIA 217  
 Db 328 ASNFYIRRALIITLVNATQCTLLCPPRRLCTDNGIMIA 364  
 QY 364 PK 364  
 RESULT 8  
 Q8JFW3 PRELIMINARY; PRT; 404 AA.  
 AC Q8JFW3;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE SI:4221013.4 (Novel glycoproteinase).  
 GN Name=dz22814.6;  
 OS Brachydanio rerio (zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NC NCB1\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Babage A.;  
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL591593; CAD3471.1; -.  
DR MEROPS; M22.004; -.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR00905; Peptidase M22.  
DR InterPro; IPR009180; Pept M22 Osa1gl.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PIRSF; PIRSF04537; Osa1glc\_ptide; 1.  
DR PRINTS; PR00789; OSIALOPTASE.  
DR ProDom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
KM Protease.  
SQ SEQUENCE 404 AA; 43956 MW; 3A6B1D050737E35 CRC64;  
Query Match 62.0%; Score 859; DB 2; Length 404;  
Best Local Similarity 62.3%; Pred. No. 3,4e-65;  
Matches 165; Conservative 38; Mismatches 62; Indels 0; Gaps 0;  
QY 1 MEAAHLLTRLTNKVPEPPLVLLISGGHCLALVQGSDFLLGKSLDIAFGMDLVKVAR 60  
DB MEAAHLLTRLMHPDPPPLVLLVSGHSLALAKGIDFLLGQTLDEBAAGTLDKIARR 195  
QY 61 LSLIKHECSTWSGKAIEHLAKOGNRFHPDIKPLHAKNCDPFTGLQHTVDKIIMK 120  
DB LSLNNHPECSTLSSGQAIERLAKSGDRLAFHISPMGQNYDCNFSFAGLRTOITGAINK 255  
QY 121 EKEEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCRKORDLPQNNALVAGGV 180  
DB EKEEGVEAGQFLSCVKDIAAASQHTVASHLAKRTTRAILFCRSKGLPDEQNPFLIVSGGV 315  
QY 181 ASNFYIRRALIETNATQCTLLCPPLRCTDNGIMIAMNGIERLRAGLILHDIGIRYE 240  
DB ASNFYIRQLKITTDATGHLCLCPSPKCTDNGVMIAMNGIERLRKQGGILSYSEVSVE 375  
QY 241 PKCPILGVDISKVEGASIVPOLKM 265  
DB PKAPLGLDITSEVKEAAIKVPLKTL 400  
RESULT 9  
Q8JFR7 PRELIMINARY; PRT; 404 AA.  
AC Q8JFR7;  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE SI:d72B14.6 (Novel glycoprotease).  
GN Name=SI:d72B14.6;  
OS Brachydanio rerio (zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_Taxid=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Clark G.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL672217; CAD3443.1; -.  
DR MEROPS; M22.004; -.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR00905; Peptidase M22.  
DR InterPro; IPR009180; Pept M22 Osa1gl.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PIRSF; PIRSF04537; Osa1glc\_ptide; 1.  
DR PRINTS; PR00789; OSIALOPTASE.  
DR ProDom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
KM Protease.

SQ SEQUENCE 404 AA; 44027 MW; 6FE98653A651860F CRC64;  
Query Match 61.7%; Score 855; DB 2; Length 404;  
Best Local Similarity 61.9%; Pred. No. 7.6e-65;  
Matches 164; Conservative 39; Mismatches 62; Indels 0; Gaps 0;  
QY 1 MEAAHLLTRLTNKVPEPPLVLLISGGHCLALVQGSDFLLGKSLDIAFGMDLVKVAR 60  
DB MEAAHLLTRLMHPDPPPLVLLVSGHSLALAKGIDFLLGQTLDEBAAGTLDKIARR 195  
QY 61 LSLIKHECSTWSGKAIEHLAKOGNRFHPDIKPLHAKNCDPFTGLQHTVDKIIMK 120  
DB LSLNNHPECSTLSSGQAIERLAKSGDRLAFHISPMGQNYDCNFSFAGLRTOITGAINK 255  
QY 121 EKEEGIEKGQILSSADIAATVQHTMACHLVKTRTRAILFCRKORDLPQNNALVAGGV 180  
DB EKEEGVEAGQFLSCVKDIAAASQHTVASHLAKRTTRAILFCRSKGLPDEQNPFLIVSGGV 315  
QY 181 ASNFYIRRALIETNATQCTLLCPPLRCTDNGIMIAMNGIERLRAGLILHDIGIRYE 240  
DB ASNFYIRQLKITTDATGHLCLCPSPKCTDNGVMIAMNGIERLRKQGGILSYSEVSVE 375  
QY 241 PKCPILGVDISKVEGASIVPOLKM 265  
DB PKAPLGLDITSEVKEAAIKVPLKTL 400  
RESULT 10  
Q7Q918 PRELIMINARY; PRT; 401 AA.  
AC Q7Q918;  
DT 01-MAR-2004 (Tremblrel. 26, Created)  
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE AGCP14990 (Fragment).  
GN Name=agCG46164; ORFNames=ENSGANG0000007922;  
OS Anopheles gambiae str. PEST.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
OX NCBI\_Taxid=180454;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=PEST;  
RC Anopheles Genome Sequencing Consortium;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AAB01008900; EAA09387.1; -.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR00905; Peptidase M22.  
DR InterPro; IPR009180; Pept M22 Osa1gl.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PIRSF; PIRSF04537; Osa1glc\_ptide; 1.  
DR PRINTS; PR00789; OSIALOPTASE.  
DR ProDom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
FT NON TER 1  
SQ SEQUENCE 401 AA; 43849 MW; CC9426723D3FD4F1 CRC64;  
Query Match 32.7%; Score 453.5; DB 2; Length 401;  
Best Local Similarity 38.4%; Pred. No. 1.9e-30;  
Matches 106; Conservative 50; Mismatches 109; Indels 11; Gaps 4;  
QY 1 MEAAHLLTRLTNKVPEPPLVLLISGGHCLALVQGSDFLLGKSLDIAFGMDLVKVAR 60  
DB 119 M0AHLAMAMTISTIPFLCLVSGHSLLVVESTARRLIGETLDDAPEGALDKIARR 178  
QY 61 LSLIKHECSTWSGKAIEHLAKOG-----NRFHPDIKPLHAKNCDPFTGLQHTVDK 115  
DB 179 LKLRVAVAKYAKQMSGGQAIERLAKGAKDTSAVBEPL--PLSKYRDCQGSFAGLRKMTATR 236



RA Ashburner M., Celniker S.B.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 a genomics perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,  
 RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu U., Berman B.P.,  
 RA Beltercourt B.R., Celniker S.B., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.B.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 systematic review.";  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RG Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RG FlyBase;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AEO03513; AAF49008.1; -  
 DR IntAct; Q9VWD6; -; CG14231.  
 DR FlyBase; FBgn0031060; CG14231.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000905; Peptidase M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialg1.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialgic\_ptclds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 DR SEQUENCE 409 AA; 45328 MW; 9797F66E7D155538 CRC64;  
 SQ  
 Query Match 30.0%; Score 415.5; DB 2; Length 409;  
 Best Local Similarity 38.2%; Pred. No. 3.6e-27;  
 Matches 100; Conservative 43; Mismatches 104; Indels 15; Gaps 6;  
 QY 1 MEAAALTRRLT--KVEPFLVLLISGGHCLALVQGVSDFLLGKSLDIAPGMDLKVYA 58  
 DB 136 MEAAALQARMHPPOIGYPLCLASGGHCOLVANGGRLLTLCQTLDDAFGEAFDKIG 195  
 QY 59 RRLSLIHPECSTWSGKAIEHLAK-QGNRPFPDIKPLHNAKNCDFSPFGLOHYTDKII 117  
 DB 196 RRLRLHLPLPRFLMNGRAIHAQALASDPLATYEPPLPAQQRNCNFSFAGIKNNFSFRI 255  
 QY 118 MKKEKEGIEKQGLISSADIAATVQHTMAGLVKTRTRALLFC--KORDLLPQNNAVLY 175  
 DB 256 RARRRARTRPDGVISNNGDFCAGLRSSVSHLMRTQRAIYCYLLPRHQLFGDPPPLIV 315  
 QY 176 ASGGVANSFYTRALEITITNAITQCTLLCPRPRLCTDNGIMTAMNGIERLRAGLGLADIE 235  
 DB 316 MSGGVANNDAIYANIEHLAAGYGCGRFSPKSKYCSDNQGMIMHGVDEL-----LDKKE 369  
 QY 236 -GTYEPKCPGLGVDSKEVGEA 256  
 DB 370 ASTRYDYD--SDIQSGAGFA 388  
 RESULT 13  
 022145 PRELIMINARY; PRT; 480 AA.  
 AC 022145; Q8VWL2;  
 DT 01-JUN-1998 (TREMBlrel. 05, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 05-JUN-2004 (TREMBlrel. 27, Last annotation update)  
 DE Putative O-sialoglycoprotein endopeptidase (Sialoglycoprotease

DE GCP1).  
 GN Name=At2g45270; Synonyms=GCP1;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCB1\_TaxID=3702;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
 RA Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,  
 RA Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Town C.D., Kaul S.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Hausenuhl K., Adamska I.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
 RA Carinci P., Chen H., Cheuk R., Hayaishiaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narueeka M., Nguyen M., Palm C.U., Sakurai T., Satou M.,  
 RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Chan M.W., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Carinci P., Chen H., Cheuk R., Hayaishiaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narueeka M., Nguyen M., Palm C.U., Sakurai T., Satou M.,  
 RA Seki M., Shim P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC002387; AAB82636.2; -  
 DR EMBL: AY024338; AAK00530.1; -  
 DR EMBL: AY063864; AAL36220.1; -  
 DR EMBL: AY117283; AAM51358.1; -  
 DR PIR; B84888; B84888.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; P:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR InterPro; IPR000905; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialg1.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialgic\_ptclds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 DR SEQUENCE 480 AA; 52995 MW; 20DD6A86ACC1FFAD CRC64;  
 SQ  
 Query Match 29.4%; Score 407.5; DB 2; Length 480;  
 Best Local Similarity 37.0%; Pred. No. 2.1e-26;  
 Matches 104; Conservative 37; Mismatches 93; Indels 47; Gaps 7;  
 QY 1 MEAAALTRRLT--NKVEPFLVLLISGGHCLALVQGVSDFLLGKSLDIAPGMDLKVYA 59  
 DB 135 MEAAALVARLVQGBISFPFMAILLISGHNLLVLANGLQCTQLGTTVDALIGAFPKTKA 254  
 QY 60 RSLIKHPECSTWSGKAIEHLAKQGNRPFPDIKPLHNAKNCDFSPFGLOHYTDKII 119  
 DB 255 WIGLDMH-----RSSGPVAVBELALSGDAKSVFNVPMKVKHCKCNFSYAGIKTQVRLAIEA 309



QY 120 KEKEGIEKGOILSSA-----ADIAATVOHTMACHLVKTRTHRAILFCORDLPPONN 171  
DB 310 KE-----IDAKCPVASATBEDRRNRADIAASFORVAVLHEEKCERADIMALE---LEPSI 362  
QY 172 AVLVASGVANFYTRALLETILNATQCTLLCPPRCLTDNGIMIANNGIERLRLAGLIL 231  
DB 363 KHWIISGVANSKYRLRLNINIENKMLKLVCPSPSLCTDNGVMAVMTGLLEHFRVG--- 418  
QY 232 HDIGIRYE-----PKCPGLVDISKEYGEA 256  
DB 419 -----RYDPPPPATPEPDYDYDLRPRMPLGBEYAKGSEA 453  
RESULT 14  
ID 073H71 PRELIMINARY; PRT; 335 AA.  
AC 073H71;  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Peptidase, M22 family protein.  
GN OrderedCusNames=WD0699;  
OC Wolbachia pipiensis wmel.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Wolbachiae; Wolbachia.  
OX NCBI\_TaxId=66077;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=15024419;  
RA Wu M., Sun L.V., Vamathevan J.J., Riegler M., DeBoy R.T.,  
RA Brownie J.C., McGraw E.A., Martin W., Esser C., Ahmedinejad N.,  
RA Wiegand C., Madupu R., Beaman M.J., Brinkac L.M., Daugherty S.C.,  
RA Durkin A.S., Kolonay J.F., Nelson W.C., Mohamoud Y., Lee P.,  
RA Berry K.J., Young W.B., Uetzelack T.R., Weidman J.F., Eisen J.A.,  
RA Paulsen I.T., Nelson K.E., Tettelin H., O'Neill S.L., Egan J.A.,  
RT "Phylogenomics of the reproductive parasite Wolbachia pipiensis wmel:  
RT a streamlined genome overrun by mobile genetic elements".  
RL Plos Biol. 2:327-341(2004).  
DR EMBL: AE017258; AAS14395.1; --  
DR TIGR; WD0699;  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR009095; Peptidase M22.  
DR InterPro: IPR009180; Pept M22\_Osialgl.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PIRSF; PIRSF004537; Osialglc\_ptds; 1.  
DR PRINTS; PR00789; OSIALOPTASB.  
DR Prodom; PD002367; Peptidase M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
DR KW Complete proteome.  
SQ SEQUENCE 335 AA; 36634 MW; F9194CDA58BC9390 CRC64;  
Query Match 26.6%; Score 369; DB 2; Length 335;  
Best Local Similarity 34.7%; Pred. No. 2.8e-23;  
Matches 84; Conservative 46; Mismatches 86; Indels 26; Gaps 5;  
QY 1 MEAAATLRLTNKVFPPVLVLLISGHCILALVQGVSPFLILGSLDIAPGMDLVKAR 60  
DB 112 LEAAVALYRLHVKPFLVLVLLISGHCQFLIADQVGYKILGERTLDLSLGAAPKVAKM 171  
QY 61 LSLIHPECSMTSGGKATLHAKOGNRFHFDIKPPLHAKNOCPSFTGLQHTVTKIMK 120  
DB 172 LGL-----SYPCGPLIEKLAKGNGTRFKLPRAMIRSGCNFSFGIKTAVKMLVQBL 224  
QY 121 EKEGIEKGOILSSAADIAATVOHTMACHLVKTRTHRAILFCORDLPPONN 180  
DB 225 KWS-----QV 273  
QY 161 ASNFYIRALLETILNATQCTLLCPPRCLTDNGIMIANNGIERLRLAGLILHDIGIRYE 240  
DB 274 AANNFLAEKLOKOHIN--LNIFFPPNDLCTNAILMVGMTGIERLOKNT-----IDPLAFA 325

QY 241 PK 242  
DB 326 PR 327  
RESULT 15  
ID 092LH8 PRELIMINARY; PRT; 360 AA.  
AC 092LH8;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).  
GN ORFNames=SMC03230;  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
OX NCBI\_TaxId=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kles B., Lelaure V., Masny D.,  
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger F.,  
RA Renard C., Thebault P., Vandenbol M., Weidner S., Gallibert F.,  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium meliloti strain 1021".  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
DR EMBL; AL591792; CAC47657.1; --  
DR GO; GO:0016787; F:hydrolyase activity; IEA.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR009095; Peptidase M22.  
DR InterPro: IPR009180; Pept M22\_Osialgl.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PIRSF; PIRSF004537; Osialglc\_ptds; 1.  
DR PRINTS; PR00789; OSIALOPTASB.  
DR Prodom; PD002367; Peptidase M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
DR KW Complete proteome; Hydrolase.  
SQ SEQUENCE 360 AA; 37906 MW; A07F946B562EA86 CRC64;  
Query Match 25.7%; Score 356.5; DB 2; Length 360;  
Best Local Similarity 35.6%; Pred. No. 3.5e-22;  
Matches 95; Conservative 38; Mismatches 91; Indels 43; Gaps 7;  
QY 1 MEAAATLRLTNKVFPPVLVLLISGHCILALVQGVSPFLILGSLDIAPGMDLVKAR 60  
DB 116 LEHALYRLHVKPFLVLVLLISGHCQFLIADQVGYKILGERTLDLSLGAAPKVAKM 175  
QY 61 LSLIHPECSMTSGGKATLHAKOGNRFHFDIKPPLHAKNOCPSFTGLQHTVTKIMK 120  
DB 176 LGL-----GGPAVRAAQAQAERFDPPLVVGARLDPSGLTAVKQAQSL 228  
QY 111 -HTVTKIMKKEGIEKGOILSSAADIAATVOHTMACHLVKTRTHRAILFCORDLPPONN 169  
DB 229 GAVTQD-----ADVCAFORAISTLDRVGRGLKRRP-ADFAV 269  
QY 170 NNAVLVASGVANFYTRALLETILNATQCTLLCPPRCLTDNGIMIANNGIERLRLAGL 229  
DB 270 DQPALVAVGVANQTLRTTQSLDEHGFRTAPLQCTDNAAAMIAWAGARLAAAG- 328  
QY 230 ILHDIGIRYEPKC--PLGVDISKEYG 254  
DB 329 ---PADGLDAAPSRWPLDSKAKALIG 352  
Search completed: February 16, 2005, 13:08:40  
Job time : 47.2739 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 08:37:51 ; Search time 48.0598 Seconds  
(without alignment)  
1923.349 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVQGVSDPFLILGKSLDI.....DISKVGSAIKVPQLKMEI 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues  
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	267	5	ABG96487 Novel hum
2	1240	100.0	414	5	ABG96478 Novel hum
3	1240	100.0	414	5	ABB05481 Human O-8
4	1240	100.0	414	6	ABJ26654 Human pro
5	1203	97.0	414	5	AAE29234 Human gly
6	1203	97.0	414	6	ABG71161 Human hum
7	1203	97.0	414	6	ABU09569 Human gly
8	960	77.4	364	6	ADA54471 Human pro
9	332.5	26.8	409	4	ABB69133 Drosophila
10	326	26.3	463	3	AAV52216 Arabidops
11	319	25.7	439	3	AAAG19287 Arabidops
12	319	25.7	439	3	AAAG19285 Arabidops
13	319	25.7	444	3	AAAG19285 Arabidops
14	257	20.7	382	6	ABU22934 Protein e
15	248.5	20.0	348	6	ABU35232 Protein e
16	247	19.9	312	6	ABU16575 Protein e
17	247	19.9	350	8	ADL05040 M. catara
18	241	19.4	343	6	ABU39303 Protein e
19	239	19.3	251	5	AAE31054 Ehrlichia
20	231	18.6	341	4	AAU36205 Pseudomon
21	231	18.6	341	6	ABU38276 Protein e
22	221	18.5	341	7	ADG73342 P aerugin
23	231	18.6	401	7	ABO68626 Pseudomon
24	229	18.5	341	7	ADG73344 P aerugin
25	224	18.1	342	3	AAV52202 Haemophil

26	224	18.1	342	4	AAU35450 Haemophil
27	224	18.1	342	6	ABU30280 Protein e
28	213.5	17.2	347	4	ABBE1112 Drosophila
29	213	17.2	337	4	AAU38187 Salmonella
30	213	17.2	337	6	ABU47561 Protein e
31	212	17.1	340	6	ABU40514 Protein e
32	212	17.1	357	7	ADP06228 Bacteri
33	210	16.9	337	6	ABU50237 Protein e
34	209	16.9	326	4	AAE96423 Putative
35	208	16.8	335	6	ABU27480 Protein e
36	207.5	16.7	341	6	ABU40069 Protein e
37	206	16.6	337	3	AAV52204 Escherich
38	206	16.6	337	4	AAU34711 E. coli c
39	206	16.6	337	7	ABU28771 Protein e
40	205	16.5	343	7	ABO62704 Klebsiella
41	201.5	16.2	421	5	ABG96491 Novel hum
42	200	16.1	325	2	AAE26325 Glycoprot
43	200	16.1	325	3	AAV52203 Pasteurel
44	199.5	16.1	350	6	ABU25741 Protein e
45	199	16.0	354	6	ABU37844 Protein e

ALIGNMENTS

RESULT 1  
ID ABG96487 standard; protein, 267 AA.  
XX  
AC ABG96487;  
XX  
DT 11-DEC-2002 (first entry)  
XX  
XX Novel human metalloprotease MPI fragment #1.  
DE  
XX  
XX Metalloprotease: MP-1; immune disorder; glutamate transport; cancer;  
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
KW neurological disorder.  
XX  
XX Homo sapiens.  
OS  
XX WO200272751-A2.  
PN  
XX 19-SEP-2002.  
PD  
XX 05-FEB-2002; 2002WO-US003353.  
PE  
XX 05-FEB-2001; 2001US-0266518P.  
PR 10-APR-2001; 2001US-0282814P.  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
XX WPI, 2002-723329/78.  
XX N-PSDB; ABS76639.  
XX  
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloprotease activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.  
XX  
XX Claim 5; Page 29; 473pp; English.  
CC The invention describes an isolated nucleic acid molecule (I) encoding a  
metalloprotease (MP-1). (I) is useful for preventing, treating, or

CC ameliorating a medical condition, particularly an immune disorder, an  
 CC aberrant glutamate transport or motor neuron disorder, such as  
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
 CC condition. The compositions and methods are also useful for diagnosing,  
 CC prognosticating, treating, ameliorating and/or treating disorders  
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
 CC sequence of a metalloprotease MP1 protein

XX Sequence 267 AA;

Query Match 100.0%; Score 1240; DB 5; Length 267;

Best Local Similarity 100.0%; Pred. No. 7,9e-131; Mismatches 0; Indels 0; Gaps 0;

Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSTMSGKAIIEHLAKQGNRF 60  
 DB 29 LIALVQGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSTMSGKAIIEHLAKQGNRF 88  
 QY 61 HFDIKPPLHAHAKNCDPFTGLQHTVDKTIIMKKEKEBEGIKGQILSSADIAATVQHTMAC 120  
 DB 89 HFDIKPPLHAHAKNCDPFTGLQHTVDKTIIMKKEKEBEGIKGQILSSADIAATVQHTMAC 148  
 QY 121 HLVKRTHRAILFCCKRDLLPNNNAVIVASGVASNFYIRRALEILTNAQCTLLCPPPRL 180  
 DB 149 HLVKRTHRAILFCCKRDLLPNNNAVIVASGVASNFYIRRALEILTNAQCTLLCPPPRL 208  
 QY 181 CTDNGIMIAMNGIERLRAGILGILHDIGIRYEPKPCPLGVDISKVEGASIKVPQLKMEI 239  
 DB 209 CTDNGIMIAMNGIERLRAGILGILHDIGIRYEPKPCPLGVDISKVEGASIKVPQLKMEI 267

RESULT 2

ABG96478 ID ABG96478 standard; protein; 414 AA.

XX ABG96478;

DT 11-DEC-2002 (first entry)

DE Novel human metalloprotease MP1.

XX Metalloprotease: MP-1; immune disorder; glutamate transport; cancer;  
 KM motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
 KM reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
 KM genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
 KM Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
 KM Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
 KM liver disease; renal disease; immune disorder; rheumatoid arthritis;  
 KM acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
 KM emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
 KM neurological disorder.

OS Homo sapiens.

PN WO200272751-A2.

PD 19-SEP-2002.

PF 05-FEB-2002; 2002WO-US003353.

XX 05-FEB-2001; 2001US-0266518P.  
 PR 10-APR-2001; 2001US-0282814P.

PA (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
 XX MPI: 2002-723329/78.  
 DR N-PSDB; ABS76635.  
 XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
 PT treating, or ameliorating diseases associated with aberrant  
 PT metalloprotease activity, e.g. immune, metabolic, inflammatory and  
 PT neurological disorders.

PS Claim 5; Fig 1A-C; 473pp; English.

CC The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC metalloprotease (MP-1). (I) is useful for preventing, treating, or  
 CC ameliorating a medical condition, particularly an immune disorder, an  
 CC aberrant glutamate transport or motor neuron disorder, such as  
 CC amyotrophic lateral sclerosis (ALS), its juvenile form or an ALS-like  
 CC condition. The compositions and methods are also useful for diagnosing,  
 CC prognosticating, treating, ameliorating and/or treating disorders  
 CC associated with MP-1 activity, e.g. diabetes, cancer, reproductive  
 CC disorders (e.g. Klinefelter's syndrome, genital warts, or germinal cell  
 CC aplasia), metabolic disorders (e.g. premature puberty, Kallman syndrome,  
 CC or Cushing's syndrome), neurodegenerative diseases (Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease or Tourette syndrome), liver  
 CC and renal diseases and immune disorders (e.g. AIDS, rheumatoid arthritis  
 CC or sepsis), pulmonary diseases (e.g. pneumonia, emphysema or cystic  
 CC fibrosis) and vascular, inflammatory and neurological disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease). This is the amino acid  
 CC sequence of a metalloprotease MP1 protein

SQ Sequence 414 AA;

Query Match 100.0%; Score 1240; DB 5; Length 414;

Best Local Similarity 100.0%; Pred. No. 1.5e-130; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSTMSGKAIIEHLAKQGNRF 60  
 DB 176 LIALVQGVSDFLILGKSLDIAPGMDLVKVARSLIKHPECSTMSGKAIIEHLAKQGNRF 235  
 QY 61 HFDIKPPLHAHAKNCDPFTGLQHTVDKTIIMKKEKEBEGIKGQILSSADIAATVQHTMAC 120  
 DB 236 HFDIKPPLHAHAKNCDPFTGLQHTVDKTIIMKKEKEBEGIKGQILSSADIAATVQHTMAC 295  
 QY 121 HLVKRTHRAILFCCKRDLLPNNNAVIVASGVASNFYIRRALEILTNAQCTLLCPPPRL 180  
 DB 296 HLVKRTHRAILFCCKRDLLPNNNAVIVASGVASNFYIRRALEILTNAQCTLLCPPPRL 355  
 QY 181 CTDNGIMIAMNGIERLRAGILGILHDIGIRYEPKPCPLGVDISKVEGASIKVPQLKMEI 239  
 DB 356 CTDNGIMIAMNGIERLRAGILGILHDIGIRYEPKPCPLGVDISKVEGASIKVPQLKMEI 414

RESULT 3

ABB05481 ID ABB05481 standard; protein; 414 AA.

XX ABB05481;

DT 19-APR-2002 (first entry).

DE Human O-6-allylglycoproteinae-like protein SEQ ID NO:2.

KM Human; O-6-allylglycoproteinae-like protein; OSGLP; enzyme.

XX Homo sapiens.

PN CN1318550-A.

PD 24-OCT-2001.

PF 19-APR-2000; 2000CN-00106834.

PR 19-APR-2000; 2000CN-00106834.  
 XX (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.  
 XX Mao Y, Xie Y;  
 PI WPI; 2002-115090/16.  
 DR N-PSDB; ABA93268.  
 XX  
 PT O-sialoglycoproteinase-like protein and encoding polynucleotide, useful  
 PT for diagnosing, preventing and treating related diseases.  
 PS Claim 1; Page 30-31 (disclosure); 38pp; Chinese.  
 CC The present sequence represents human O-sialoglycoproteinase-like protein  
 CC (OSGPRP). The present invention also describes: (1) the preparation of  
 CC the OSGPRP protein; (2) applying the OSGPRP protein in diagnosis; (3) the  
 CC prevention and/or treatment of related diseases; (4) utilizing the OSGPRP  
 CC protein in screening its agonist, excitomotor and inhibitor and preparing  
 CC an antibody against the OSGPRP protein; and (5) the use of the OSGPRP  
 CC polynucleotide sequences, proteins, agonists, excitomotors, inhibitors  
 CC and antibodies in treating diseases related to the abnormal OSGPRP gene  
 CC and in preparing the medicine composite for the treatment  
 XX Sequence 414 AA;  
 SO  
 Query Match 100.0%; Score 1240; DB 5; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-130;  
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LLAIVQGVSDPFLILGKSLDIAPGMDLVYARRLSLTKHPECSTWGGKAIIEHLAKGNRF 60  
 DB 176 LLAIVQGVSDPFLILGKSLDIAPGMDLVYARRLSLTKHPECSTWGGKAIIEHLAKGNRF 235  
 QY 61 HFDIKPPLHAKNCDPSFTGLQHTVDKILMKKEKEGIEKGQILSSAADIAATVQHTMAC 120  
 DB 236 HFDIKPPLHAKNCDPSFTGLQHTVDKILMKKEKEGIEKGQILSSAADIAATVQHTMAC 295  
 QY 121 HLKRTTRAILFCQKRDILPQNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPLRL 180  
 DB 296 HLKRTTRAILFCQKRDILPQNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPLRL 355  
 QY 181 CTDNGIMIAWNGIRLRAGILGILHDIGIRYRKPCLGVDISKVGEASIKVPQLKMEI 239  
 DB 356 CTDNGIMIAWNGIRLRAGILGILHDIGIRYRKPCLGVDISKVGEASIKVPQLKMEI 414  
 RESULT 4  
 ABJ26654  
 ID ABJ26654 standard; protein; 414 AA.  
 XX  
 AC ABJ26654;  
 XX  
 DT 01-MAY-2003 (first entry)  
 XX  
 DE Human protein modification + maintenance molecule protein SEQ ID No 8.  
 XX  
 XX Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; anticonvulsant;  
 XX cerebroprotective; antiparkinsonian; nootropic; antiinflammatory;  
 XX antitumor; hepatotropic; gynecological; antibacterial; vinicide;  
 XX protozoacide; antiparasitic; cell proliferative disease; PMOD;  
 XX protein modification and maintenance molecule; immunogenic fragment;  
 XX cancer; autoimmune; inflammatory disease; neurological disorder;  
 XX gastrointestinal; developmental; vesicle trafficking disorder; infection;  
 XX protein-protein interaction; drug-target interaction;  
 XX gene expression profile; human.  
 OS Homo sapiens.  
 XX  
 XX NC03030664-1-12.  
 XX  
 XX 03-JAN-2003.  
 PD  
 XX

PF 18-JUN-2002; 2002MO-US019360.  
 XX  
 XX 22-JUN-2001; 2001US-0300508P.  
 PR 06-JUL-2001; 2001US-0301445P.  
 PR 13-JUL-2001; 2001US-0305405P.  
 PR 09-AUG-2001; 2001US-0311442P.  
 PR 24-AUG-2001; 2001US-0314821P.  
 PR 29-AUG-2001; 2001US-0315992P.  
 PR 03-MAY-2002; 2002US-0378205P.  
 XX  
 PA (INCYTE) INCYTE GENOMICS INC.  
 XX  
 PI Gandhi AR, Kable AE, Swarnakar A, Hafalia AJA, Tran B, Duggan BW,  
 PI Warren BA, Ison CH, Honchelli CD, Nguyen DB, Lu DAM, Lee EA, Yue H;  
 PI Forsythe JF, Barroso I, Ramkumar J, Griffin JB, Li JX, Yang J;  
 PI Thangavelu K, Gietzen KJ, Ding L, Baughn MR, Borowsky ML, Yao MG,  
 PI Walla NR, Mason PM, Gururajan R, Lee S, Becha SD, Lee SY, Tran UK,  
 PI Elliott VS, Luo W, Sprague WW, Tang YT, Lu Y, Zebajadian Y;  
 DR WPI; 2003-184039/18.  
 DR N-PSDB; ABR23207.  
 XX  
 PT New isolated human PMOD polypeptide and polynucleotide, useful for  
 PT diagnosing, treating and preventing diseases or conditions associated  
 PT with the aberrant PMOD expression, e.g. cancer, AIDS, atherosclerosis and  
 PT infections.  
 PS Claim 63; Page 182-183; 225pp; English.  
 XX  
 XX The invention relates to an isolated polypeptide comprising: any of 28  
 XX sequences of 48-1256 amino acids; a natural amino acid sequence at least  
 XX 90% identical to the 28 amino acid sequences, 94% identical to a sequence  
 XX of 703 or 267 amino acids, 96% identical to a sequence of 414 amino  
 XX acids, or 97% identical to a sequence of 242 amino acids, all given in  
 XX the specification; or a biologically active or immunogenic fragment of  
 XX the isolated polypeptide. The polypeptides and polynucleotides are useful  
 XX in diagnosing, treating and preventing diseases or conditions associated  
 XX with the decreased expression of protein modification and maintenance  
 XX molecules (PMOD), such as cell proliferative diseases (e.g. cancer,  
 XX atherosclerosis), autoimmune/inflammatory diseases (e.g. AIDS,  
 XX allergies), neurological disorders (e.g. stroke, Parkinson's disease,  
 XX epilepsy), gastrointestinal (e.g. ulcer, cirrhosis), reproductive (e.g.  
 XX endometriosis), developmental, vesicle trafficking disorders, and  
 XX infections (e.g. bacterial, viral, parasitic, protozoal). These are also  
 XX useful in assessing the effects of exogenous compounds on the expression  
 XX of nucleic acid and amino acid sequences of PMOD. The PMOD or its  
 XX fragments are useful in screening compounds for effectiveness as agonist  
 XX or antagonist of the polypeptides, or in altering the expression of the  
 XX target polynucleotide and compounds that specifically bind to or modulate  
 XX the activity of the polypeptide. The microarray is useful in monitoring  
 XX or measuring protein-protein interactions, drug-target interactions, and  
 XX gene expression profiles. This sequence represents a human PMOD protein  
 XX of the invention  
 SO Sequence 414 AA;  
 Query Match 100.0%; Score 1240; DB 6; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-130;  
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LLAIVQGVSDPFLILGKSLDIAPGMDLVYARRLSLTKHPECSTWGGKAIIEHLAKGNRF 60  
 DB 176 LLAIVQGVSDPFLILGKSLDIAPGMDLVYARRLSLTKHPECSTWGGKAIIEHLAKGNRF 235  
 QY 61 HFDIKPPLHAKNCDPSFTGLQHTVDKILMKKEKEGIEKGQILSSAADIAATVQHTMAC 120  
 DB 236 HFDIKPPLHAKNCDPSFTGLQHTVDKILMKKEKEGIEKGQILSSAADIAATVQHTMAC 295  
 QY 121 HLKRTTRAILFCQKRDILPQNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPLRL 180  
 DB 296 HLKRTTRAILFCQKRDILPQNNAVLVASGVASNFYIRRALEILTNAQCTLLCPPLRL 355  
 QY 181 CTDNGIMIAWNGIRLRAGILGILHDIGIRYRKPCLGVDISKVGEASIKVPQLKMEI 239

Db 356 CTDNGIMIAMNGIERLRAGILHDIEGIRYBPKPLGVDISKVEASIKVQLKMEI 414

RESULT 5

ID AAE29234 standard; protein; 414 AA.

XX AAE29234;

XX 27-JAN-2003 (first entry)

XX Human glycoprotease 28472 protein.

XX Human; adenosine deaminase; seven transmembrane domain receptor; cancer;  
 KW 7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arrhythmia;  
 KW rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;  
 KW hyperextension; ischaemic heart disease; obesity; myocardial infarction;  
 KW endothelial cell disorder; Grave's disease; psoriasis; brain disorder;  
 KW Parkinson's disease; Alzheimer's disease; haematopoietic disorder;  
 KW cerebral oedema; metabolic disorder; liver disorder; platelet disorder;  
 KW chromosome mapping; tissue typing; gene therapy; neuroprotective;  
 KW cytosolic; anorectic; cardiac; haemostatic.

XX Homo sapiens.

XX Location/Qualifiers

FT 1..108 /note= "Non-transmembrane domain; N-terminal cytoplasmic

FT domain"

FT /note= "Endopeptidase O-sialoglycoprotein hydrolase

FT metalloprotease zinc glycoprotease sialoglycoprotease

FT domain"

FT /note= "Transmembrane domain"

FT /note= "Non-transmembrane domain; non-cytoplasmic loop"

FT /note= "Glycoprotease domain"

FT /note= "Transmembrane domain"

FT /note= "Transmembrane domain"

FT /note= "Non-transmembrane domain; cytoplasmic domain"

FT /note= "Transmembrane domain"

FT /note= "Non-transmembrane domain"

FT /note= "Sialoglycoprotease type domain"

FT /note= "Transmembrane domain"

FT /note= "Non-transmembrane domain"

FT /note= "Transmembrane domain"

FT /note= "Non-transmembrane domain"

FT /note= "Transmembrane domain"

FT /note= "Non-transmembrane domain"

FT /note= "Transmembrane domain"

FT /note= "Non-transmembrane domain"

FT /note= "Transmembrane domain"

PS Claim 1; Fig 8; 178pp; English.

XX The present invention relates to novel 38650, 28472, 5495, 65507, 81588  
 CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-  
 CC protease or seven transmembrane domain (7TM) receptor family members.  
 CC Sequences of the invention are useful in diagnosing and treating cancer  
 CC or aberrant cellular proliferation and/or differentiation (e.g. colon or  
 CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid  
 CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,  
 CC hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,  
 CC myocardial infarction, thrombus) including endothelial cell disorders  
 CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain  
 CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),  
 CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet  
 CC disorders. They are also useful in screening assays, predictive medicine  
 CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials  
 CC and pharmacogenetics) and prophylactic and therapeutic methods. The  
 CC nucleic acids may also be used in chromosome mapping, tissue typing and  
 CC forensic biology and as surrogate markers. Sequences of the invention are  
 CC also used in gene therapy. The present sequence is human glycoprotease  
 CC 28472 protein

XX Sequence 414 AA;

Query Match 97.0%; Score 1203; DB 5; Length 414;  
 Best Local Similarity 97.1%; Pred. No. 2,3e-126;  
 Matches 222; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LIALVGVSDPFLILGKSLDIAPGMDLVARRSLIKHPGCTSGGKAIETHLAKQGNRF 60

DB 176 LIALVGVSDPFLILGKSLDIAPGMDLVARRSLIKHPGCTSGGKAIETHLAKQGNRF 235

QY 61 HPDIKPRPLHANPCPSFGLQHTYDKITMKKEKEGIRKQGLSSAADIAAVQHTMAC 120

DB 236 HPDIKPRPLHANPCPSFGLQHTYDKITMKKEKEGIRKQGLSSAADIAAVQHTMAC 295

QY 121 HLVKRTRAILFCRQDRLPNNNAVVASGVASNYIRRALIETLNATQCTLLCPPEPL 180

DB 296 HLVKRTRAILFCRQDRLPNNNAVVASGVASNYIRRALIETLNATQCTLLCPPEPL 355

QY 181 CTDNGIMIAMNGIERLRAGILHDIEGIRYBPKPLGVDISKVEASIKVQLKMEI 239

DB 356 CTDNGIMIAMNGIERLRAGILHDIEGIRYBPKPLGVDISKVEASIKVQLKMEI 414

RESULT 6  
 ABG71161 standard; protein; 414 AA.

XX ABG71161;

XX 30-JAN-2003 (first entry)

XX Novel human glycoprotease 28472.

XX Cancer; aberrant cell proliferation; aberrant cell differentiation;  
 KW breast cancer; ovarian cancer; prostate cancer; colon cancer;  
 KW lung cancer; immune disorder; heart disorder; cardiovascular disorder;  
 KW endothelial disorder; hematopoietic disorder; blood vessel disorder;  
 KW brain disorder; pain; metabolic disorder; liver disorder; diabetes;  
 KW platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;  
 KW autoimmune disorder; hypertension; atherosclerosis; heart failure;  
 KW myocardial infarction; ischaemic heart disease; Croun's disease;  
 KW Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;  
 KW cerebral ischaemia; peripheral neuropathy; Alzheimer's disease;  
 KW Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; enzyme.

XX Homo sapiens.

XX Key Location/Qualifiers

FT 138..152 /label= Glycoprotease\_domain

FT Domain

PN WO200277233-A2.  
 XX 03-OCT-2002.  
 PD  
 XX 08-NOV-2001; 2001WO-US046724.  
 PF  
 XX 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 XX Leiby KR, Kapeller-Libermann R, Glucksmann M;  
 PI  
 DR WPI: 2003-029338/02.  
 DR N-PSDB; ABS57020.  
 XX  
 PT New adenosine deaminase, glycoprotease and seven transmembrane domain  
 PT nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,  
 PT 81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or  
 PT hypertension.  
 XX  
 PS Claim 4; Fig 8A-B; 178pp; English.  
 XX  
 CC The invention describes isolated 38650, 28472, 5495, 65507, 81588 and  
 CC 14354 nucleic acid molecules (I) and their encoded polypeptides (II). The  
 CC 38650 nucleic acid molecule comprises a sequence encoding adenosine  
 CC deaminase. The 28472 nucleic acid molecule comprises a sequence encoding  
 CC a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise  
 CC sequences that encode a human seven transmembrane domain (7TM). The  
 CC 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide  
 CC sequences are useful for diagnosing, preventing or treating a subject  
 CC with or at risk of developing a disorder, e.g. cancer or aberrant  
 CC cellular proliferation and/or differentiation (e.g. breast, ovarian,  
 CC prostate, colon or lung cancer), immune disorders, heart disorders,  
 CC cardiovascular disorders, endothelial disorders, hematopoietic disorders,  
 CC blood vessel disorders, brain disorders, pain and metabolic disorders,  
 CC liver disorders or platelet disorders. These disorders include carcinoma,  
 CC sarcoma, leukaemia, Hodgkin's disease, autoimmune disorders,  
 CC hypertension, atherosclerosis, heart failure, myocardial infarction,  
 CC ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki  
 CC syndrome, Raynaud's disease, aneurysm, cerebral ischaemia, peripheral  
 CC neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,  
 CC cachexia or diabetes. This is the amino acid sequence of the novel human  
 CC glycoprotease 28472  
 CC  
 XX Sequence 414 AA:  
 SQ  
 Query Match 97.0%; Score 1203; DB 6; Length 414;  
 Best Local Similarity 97.1%; Pred. No. 2.3e-126;  
 Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

AC ABU09569;  
 XX  
 DT 08-JUL-2003 (first entry)  
 DT  
 XX  
 XX Human glycoprotease encoded by cDNA 28472.  
 DE  
 XX  
 XX Human; enzyme; cancer; aberrant cellular proliferation; differentiation;  
 KW immune disorders; heart disorder; brain disorder;  
 KW cardiovascular disorder; endothelial cell disorder; pain disorder;  
 KW haematopoietic disorder; blood vessel disorder; metabolic disorder;  
 KW liver disorder; platelet disorder; glycoprotease.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX US2003009017-A1.  
 PN  
 XX  
 PD 09-JAN-2003.  
 PD  
 XX  
 XX 08-NOV-2001; 2001US-00012140.  
 PF  
 XX  
 XX 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX  
 XX (LEIBY) LEIBY K R.  
 PA (KAP) KAPPELLER-LIBERMANN R.  
 PA (GLUC) GLUCKSMANN M A.  
 XX  
 XX Leiby KR, Kapeller-Libermann R, Glucksmann MA;  
 PI  
 DR WPI: 2003-428888/40.  
 DR N-PSDB; ACA60887.  
 XX  
 CC New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid  
 CC molecules, useful for diagnosing, treating cancer, pain, or immune,  
 CC heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic  
 CC and liver disorders.  
 XX  
 PS Claim 4; Fig 8; 90pp; English.  
 XX  
 CC The invention relates to an isolated 38650 (encoding adenosine  
 CC deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7  
 CC transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or  
 CC a sequence which is at least 60% identical to the six nucleic acids or  
 CC their open reading frames, fragments of at least 15 nucleotides,  
 CC naturally occurring variants, or a DNA insert of the plasmid deposited  
 CC with the American Type Culture Collection as Accession No. not defined in  
 CC the specification, which encodes the amino acid sequence). Also included  
 CC are a host cell containing the nucleic acids (used to produce the  
 CC proteins), the encoded proteins, an antibody that selectively binds to  
 CC the polypeptide, and identifying a compound that binds to/modulates the  
 CC activity of the polypeptide. The nucleic acid molecules, polypeptides and  
 CC methods are useful for diagnosing, treating cancer, aberrant cellular  
 CC proliferation and/or differentiation, immune disorders, heart disorders,  
 CC cardiovascular disorders including endothelial cell disorders,  
 CC haematopoietic disorders, blood vessel disorders, brain disorders, pain  
 CC and metabolic disorders, liver disorders and platelet disorders (many  
 CC examples of these disorders are given in the specification). The present  
 CC sequence is the Human glycoprotease encoded by cDNA 28472  
 CC  
 XX Sequence 414 AA:  
 SQ  
 Query Match 97.0%; Score 1203; DB 6; Length 414;  
 Best Local Similarity 97.1%; Pred. No. 2.3e-126;  
 Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

QY      121  HLVRTRRAILFCQKORDLLPQNNNAVLVAGGVASNFYIRRALEILTNAQCTLLCPPPRL 180
DB      296  HLVRTRRAILFCQKORDLLPQNNNAVLVAGGVASNFYIRRALEILTNAQCTLLCPPPRL 355
QY      181  CTDNGIMIAMNGIERLRAGLGIHDIESGIRYBPKPLGVDISKVEGASIKVPOLKMEI 239
DB      356  CTDNGIMIAMNGIERLRAGLGIHDIESGIRYBPKPLGVDISKVEGASIKVPOLKMEI 414

RESULT 8
ID      ADA54471  standard; protein; 364 AA.
AC      ADA54471;
XX
XX      20-NOV-2003  (first entry)
DT
XX
XX      Human protein, SEQ ID 2039.
DE
XX
XX      Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KM      Gene Therapy; human; secretory protein; membrane proteins; cancer;
KM      Inflammatory disease; osteoporosis; neurological disease.
XX
XX      Homo sapiens.
OS
XX      EPI293569-A2.
PN
XX      19-MAR-2003.
PD
XX      21-MAR-2002; 2002EP-00006586.
PF
XX      14-SEP-2001; 2001JP-00328381.
PR      24-JAN-2002; 2002US-0350435P.
XX
XX      (HELI-) HELIX RES INST.
PA      (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI      Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
PI      Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
PI      Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
XX      WPI; 2003-395539/38.
DR      N-PSDB; ADA52832.
XX
XX      New polynucleotides encoding full-length polypeptides, e.g. secretory
PT      and/or membrane proteins, useful for developing medicines for diseases in
PT      which the gene is involved, or as target molecules for gene therapy.
XX
PS      Claim 14; SEQ ID NO 2039; 205BP; English.
XX
XX      The present invention relates to novel human secretory or membrane
CC      proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC      ADA54071). The coding sequences are useful in the gene therapy of
CC      diseases caused by abnormalities of the proteins, e.g. cancer,
CC      inflammatory diseases, osteoporosis or neurological disease.
XX
XX      Sequence 364 AA;
SQ

Query Match      77.4%; Score 960; DB 6; Length 364;
Best Local Similarity 98.4%; Pred. No. 5e-99;
Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

DB      296  HLVRTRRAILFCQKORDLLPQNNNAVLVAGGVASNFYIRRALEILTNAQCTLLCPPPRL 355
QY      181  CTDNGIMIA 189
DB      356  CTDNGIMIA 364

RESULT 9
ID      ABB69133  standard; protein; 409 AA.
AC      ABB69133;
XX
XX      26-MAR-2002  (first entry)
DT
XX
XX      Drosophila melanogaster polypeptide SEQ ID NO 34191.
DE
XX      Drosophila; developmental biology; cell signalling; insecticide;
KM      pharmaceutical.
XX
XX      Drosophila melanogaster.
OS
XX      WO200171042-A2.
PN
XX      27-SEP-2001.
PD
XX      23-MAR-2001; 2001WO-US009231.
PF
XX      23-MAR-2000; 2000US-0191637P.
PR      11-JUL-2000; 2000US-00614150.
XX
XX      (PEKE ) PE CORP NY.
PA
XX      Venter JC, Adams M, Li PWD, Myers EW;
PI
XX      WPI; 2001-656860/75.
DR      N-PSDB; ABL13236.
XX
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
PT      genes from Drosophila and for elucidating cell signalling and cell-cell
PT      interactions.
XX
XX      Disclosure; SEQ ID NO 34191; 21bp + Sequence Listing; English.
PS
XX
XX      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC      sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC      ABB72072). The sequence data for this patent did not form part of the
CC      printed specification, but was obtained in electronic format directly
CC      from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX      Sequence 409 AA;
SQ

Query Match      26.8%; Score 332.5; DB 4; Length 409;
Best Local Similarity 35.9%; Pred. No. 3.9e-28;
Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;

```

Qy 179 RLCTDNGIMIAMNGIERLRAGILHDIIE-GIRYEPKCPGLVDISKVEGEA 228  
Db 347 RYCSDNGVMIAMHGVGEOL-----LQDKREASTRYDYD--SIDIGSAGFA 388

RESULT 10  
ID AAY52216 standard; protein; 463 AA.  
XX AAY52216;  
XX AAY52216;  
XX 09-FEB-2000 (first entry)  
XX Arabidopsis thaliana yjD protein homologue.  
XX  
XX yjD protein; essential; Gram positive; Gram negative; conserved; motif;  
XX identification; antagonist; antibacterial; antibiotic; broad spectrum;  
XX treatment; infection; resistance; drug target.  
XX Arabidopsis thaliana.  
XX  
XX Key Location/Qualifiers  
XX Region 86..96  
XX /note= "yjD conserved motif 3"  
XX Region 111..131  
XX /note= "yjD conserved motif 4"  
XX Region 152..198  
XX /note= "yjD conserved motif 2"  
XX Region 208..259  
XX /note= "yjD conserved motif 1"  
XX  
XX W0954470-A2.  
XX  
XX 28-OCT-1999.  
XX  
XX 20-APR-1999; 99WO-EP002635.  
XX  
XX 22-APR-1998; 98GB-00008423.  
XX  
XX (GLAX ) GLAXO GROUP LTD.  
XX  
XX Arigoni F, Edgerton MD, Loferer H, Peltesch MC;  
XX WPI; 2000-013253/01.  
XX  
XX Novel bacterial polypeptides used to identify broad spectrum antibiotics.  
XX  
XX Claim 1; Fig 1; 55pp; English.  
XX  
XX Proteins AAY52202-Y52217 encompass a novel family of proteins designated  
XX the yjD family, after the name given to the *Escherichia coli* family  
XX member. These proteins are essential for the survival of both Gram  
XX negative and Gram positive bacteria, although no function has as yet been  
XX ascribed to these proteins. The yjD proteins, fragments of yjD proteins  
XX (for example, fragments encompassing one or more conserved yjD motifs  
XX such as AAY52218-Y52288) and nucleotides encoding them can be used to  
XX identify antagonists and broad spectrum antibacterial compounds. These  
XX antagonists and compounds can be used to treat a wide range of bacterial  
XX infections. New antibiotics are urgently needed, as serious bacterial  
XX infections and antibiotic resistant strains are becoming increasingly  
XX prevalent. The proteins of the invention are essential proteins for  
XX bacterial viability, and represent new targets for antibiotics  
XX  
XX Sequence 463 AA;  
XX

Query Match 26.3%; Score 326; DB 3; Length 463;  
Best Local Similarity 33.6%; Pred. No. 2.6e-27;  
Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;  
Qy 1 LILVGVSVDFILAGNSLIDIAFGLMDKVARRLSLIKPECSWGGRAIEHLAKQNR 60  
Db 221 LILVLAHLGCGYTOGLTIVDDAIGFAFDKATKGLDMH-----RSGRAVERLEALGDAK 275

Qy 61 HFDIKPRLHMANCDPSTGLQHTVDKILMKKEKEGIEKGQILSSAADIAATVQHTMAC 120  
Db 276 SVKFNVPKMXCHDCNFSYAGLKTQVRLAIEAKE-----IRNRADIAASFQVAVL 325

Qy 121 HLVKRTIRAILPCKQRDLIPONNAVLVAGGVASNPYIRRALEITLNATCITLLCPPL 180  
Db 326 HLEEKCEKRAIDWALE---LEPSIKHNVISGVASNKYVRLNINVENKQLKLVCPPEST 382

Qy 181 CTDNGIMIAMNGIERLRAGILHDIIEGIRYE-----PKCPGLVDISKVE 224  
Db 383 CTDNGVMVAMTGLHERVQ-----RYDPPPTBEDPYVYDLRPRMPLGEEYAKG 432

Qy 225 VGEA 228  
Db 433 RSEA 436

RESULT 11  
AAG19287  
ID AAG19287 standard; protein; 245 AA.  
XX AAG19287;  
XX  
XX 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 21031.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
XX  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0125264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127462P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 28-APR-1999; 99US-0130891P.  
XX 30-APR-1999; 99US-0131449P.  
XX 30-APR-1999; 99US-0132048P.  
XX 04-MAY-1999; 99US-0132407P.  
XX 04-MAY-1999; 99US-0132484P.  
XX 05-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 07-MAY-1999; 99US-0132487P.  
XX 07-MAY-1999; 99US-0132863P.  
XX 11-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134321P.  
XX 14-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0135621P.  
XX 27-MAY-1999; 99US-0136392P.



PR 28-MAY-1999; 99US-0116782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 21-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.

PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149930P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150565P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 18-OCT-1999; 99US-0159538P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.



PR	02-AUG-1999	99US-01465886P
PR	02-AUG-1999	99US-01465888P
PR	02-AUG-1999	99US-01465892P
PR	03-AUG-1999	99US-01470380P
PR	04-AUG-1999	99US-01472040P
PR	04-AUG-1999	99US-01473020P
PR	04-AUG-1999	99US-01473192P
PR	05-AUG-1999	99US-01477520P
PR	05-AUG-1999	99US-01472600P
PR	06-AUG-1999	99US-01473030P
PR	06-AUG-1999	99US-01474160P
PR	09-AUG-1999	99US-01474930P
PR	09-AUG-1999	99US-01479350P
PR	10-AUG-1999	99US-01481710P
PR	11-AUG-1999	99US-01483110P
PR	12-AUG-1999	99US-01483410P
PR	13-AUG-1999	99US-01485650P
PR	13-AUG-1999	99US-01486840P
PR	16-AUG-1999	99US-01493670P
PR	17-AUG-1999	99US-01491750P
PR	18-AUG-1999	99US-01494260P
PR	25-AUG-1999	99US-01497220P
PR	20-AUG-1999	99US-01497230P
PR	27-AUG-1999	99US-01510650P
PR	27-AUG-1999	99US-01510660P
PR	27-AUG-1999	99US-01510800P
PR	30-AUG-1999	99US-01510300P
PR	31-AUG-1999	99US-01514380P
PR	01-SEP-1999	99US-01519300P
PR	07-SEP-1999	99US-01523630P
PR	10-SEP-1999	99US-01530700P
PR	13-SEP-1999	99US-01537580P
PR	15-SEP-1999	99US-01540180P
PR	16-SEP-1999	99US-01540730P
PR	20-SEP-1999	99US-01545770P
PR	22-SEP-1999	99US-01551330P
PR	23-SEP-1999	99US-01554580P
PR	24-SEP-1999	99US-01556550P
PR	28-SEP-1999	99US-01564580P
PR	29-SEP-1999	99US-01565690P
PR	04-OCT-1999	99US-01571170P
PR	05-OCT-1999	99US-01577530P
PR	06-OCT-1999	99US-01578650P
PR	07-OCT-1999	99US-01580290P
PR	08-OCT-1999	99US-01582320P
PR	12-OCT-1999	99US-01583690P
PR	13-OCT-1999	99US-01592630P
PR	13-OCT-1999	99US-01592940P
PR	13-OCT-1999	99US-01593290P
PR	14-OCT-1999	99US-01593300P
PR	14-OCT-1999	99US-01593310P
PR	14-OCT-1999	99US-01596370P
PR	14-OCT-1999	99US-01596380P
PR	18-OCT-1999	99US-01595840P
PR	21-OCT-1999	99US-01607410P
PR	21-OCT-1999	99US-01607670P
PR	21-OCT-1999	99US-01607680P
PR	21-OCT-1999	99US-01607700P
PR	21-OCT-1999	99US-01610440P
PR	21-OCT-1999	99US-01608150P
PR	22-OCT-1999	99US-01608500P
PR	22-OCT-1999	99US-01609810P
PR	22-OCT-1999	99US-01609890P
PR	25-OCT-1999	99US-01614040P
PR	25-OCT-1999	99US-01614050P
PR	26-OCT-1999	99US-01614060P
PR	26-OCT-1999	99US-01613590P
PR	26-OCT-1999	99US-01613600P

Query Match	25.7%	Score 319;	DB 3;	Length 439;
Best Local Similarity	35.24;	Pred. No. 1.5e-26;		
Matches	81;	Conservative	31;	Mismatches 82;
			Indels	30;
			Gaps	5
Qy	1	L L A L V G S P F L I G K S L D I A P G D M L D K V A R R L S I K I P E C S T M S G G K A I E H L A K Q N R F	60	
Db	224	L I V L A H L D G V T Q T G T T V D D I A I G A F K T A K M L G I D N H - - - - X S G G P A V E L A L E G D A K	278	
Qy	61	H F D I K P L P L H A K N C D F S P F T G I Q H V T D K I I M K E K E B G I E K Q I I S S A - - - - - A D I P A	112	
Db	279	S V K E N V A P M K Y H K D C N F S Y A G I K T Q V R L A I E A K E - - - - I D A C K P V S S A T N E D R R N R A D I A A	334	
Qy	113	T V Q T T M A C H L V K R T H R A I L P C K Q R D L P Q N N A V I V A S G S V A S N P F I R A L E I L T M A T Q C T	172	
Db	335	S F Q V A V A L H E E K E R A I D W A L E - - - L E P S I K H V I S G S V A S N K Y R L R L A N N I V E N K O I K	391	
Qy	173	L L C P P R L C T D N G I M I A M N G I E R L R A G I G I L H D I E G I R Y E P K C P	216	
Db	392	L V C P P S L C T D N G V M A M T G L H E R V G - - - - - R Y D P P P	425	
RESULT 13				
AAG19285				
XX	ID	AAG19285	standard; protein; 444 AA.	
XX	AC	AAG19285;		
XX	DT	17-OCT-2000	(first entry)	
XX	DE	Arabidopsis thaliana	protein fragment SRQ ID NO: 21029.	
XX	KW	Protein identification; signal transduction pathway; metabolic pathway;		
XX	KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
XX	KW	termination sequence.		
XX	OS	Arabidopsis thaliana.		
XX	XX	EP1033405-A2.		
XX	XX	06-SEP-2000.		
XX	PD			
XX	PF	25-FEB-2000;	2000EP-00301439.	
XX	XX	25-FEB-1999;	99US-0121825P.	
PR	05-MAR-1999;	99US-0123180P.		
PR	09-MAR-1999;	99US-0123548P.		
PR	23-MAR-1999;	99US-0125788P.		
PR	25-MAR-1999;	99US-0126264P.		
PR	29-MAR-1999;	99US-0126785P.		
PR	01-APR-1999;	99US-0127462P.		
PR	06-APR-1999;	99US-0128234P.		
PR	08-APR-1999;	99US-0128714P.		
PR	16-APR-1999;	99US-0129845P.		
PR	19-APR-1999;	99US-0130077P.		
PR	21-APR-1999;	99US-0130449P.		
PR	23-APR-1999;	99US-0130510P.		
PR	23-APR-1999;	99US-0130891P.		
PR	28-APR-1999;	99US-0131449P.		
PR	30-APR-1999;	99US-0132048P.		
PR	30-APR-1999;	99US-0132407P.		
PR	04-MAY-1999;	99US-0132484P.		
PR	05-MAY-1999;	99US-0132485P.		
PR	06-MAY-1999;	99US-0132486P.		
PR	07-MAY-1999;	99US-0132487P.		
PR	11-MAY-1999;	99US-0132863P.		
PR	11-MAY-1999;	99US-0134256P.		
PR	14-MAY-1999;	99US-0134218P.		

PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.

PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0158293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.

PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161922P.  
 PR 28-OCT-1999; 99US-0161932P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 25.7%; Score 319; DB 3; Length 444;  
 Best Local Similarity 36.2%; Pred. No. 1.5e-26;

Matches 81; Conservative 31; Mismatches 82; Indels 30; Gaps 5;

QY 1 LIALVGVSDPFLILGKSLDIPAGMDLVKVARRLSLIKPECSMTSGGKAIENHAKOGRF 60  
 DB 229 LVLIAHKKIGQYTOGTVDVDAIGFAFDYAKWLGIDMH-----RSGGPAVEBLALEGPAK 283  
 QY 61 HFDIKPPLHAKNCDPFTGLQHVTDKTIIMKKEKEGIEKGQILSSA-----ADIAA 112  
 DB 284 SVKRNVPFKYKDKCNFSTAGLKTQVRLAIEAKE-----IDAKCPVSSATNEDRRNRDIAA 339  
 QY 113 TVQHTMACLVKRTTRAILFCQKRDLPQNNAVLVASGVASNFYIRRALBITNATQCT 172  
 DB 340 SFQVAVVLEHEKEKRAIDWALE---LEPSIKHWVISGGSVANSKTVRLINNIYENKOLX 396  
 QY 173 LACPPRLCTDNGMIAMNGIERLPAIGILHDEIGIRYBKP 216  
 DB 397 LVCPPPSLCTDNGVAVWMTGLEHFRVG-----RYDPPPP 430

## RESULT 14

ABU22934  
 ID ABU22934 standard; protein; 382 AA.

XX AC ABU22934;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #8461.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Bordetella pertussis.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 PI WPI; 2003-029926/02.  
 DR N-PSDB; ACA26804.

PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 50858; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-regulated gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 382 AA;

Query Match 20.7%; Score 257; DB 6; Length 382;

Best Local Similarity 36.1%; Pred. No. 1.2e-19;

Matches 74; Conservative 33; Mismatches 78; Indels 20; Gaps 7;

QY 2 LIALVGVSDPFLILGKSLDIPAGMDLVKVARRLSLIKPECSMTSGGKAIENHAKOGRH 61  
 DB 178 LMLVGVGKRYELGSLTLDPAAGEAPDKSKLMGL-GYP-----GGPALARLAEQDDASR 230  
 QY 62 PDIKPPLHAKNCDPFTGLQ-HVTDKTIIMKKEKEGIEKGQILSSAADIAATVQHTMAC 120  
 DB 231 YDLPRPMLHSGDLDSFSGLKTAVLTRV--KATRDGSELG--QDRADLAATQAAIYB 286  
 QY 121 HLVKRTTRAILFCQKRDLPQNNAVLVASGVASNFYIRRALBITNATQCTLLCPPRL 180  
 DB 287 VLAARAIKRL--KQTGL-----RLVVAGVGAVNALLRAHLARALKPRAEAYFPPLSL 338  
 QY 181 CTDNGMIAMNGIERLPAIGILHDEIGIRYBKP 205  
 DB 339 CTDNGMIAMFAAEHVRKAGLADIRE 363

## RESULT 15

ABU35232  
 ID ABU35232 standard; protein; 348 AA.

XX AC ABU35232;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #20759.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Moraxella catarrhalis.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyekind JW;  
PI Wali D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.

DR N-PSDB; ACA39102.

PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 63156; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway of  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 348 AA;

Query Match 20.0%; Score 248.5; DB 6; Length 348;

Best Local Similarity 33.2%; Pred. No. 9.4e-19;

Matches 66; Conservative 33; Mismatches 79; Indels 21; Gaps 5;

QY 1 LIALVGVSDPLLGKSLDIAPGMLDVARRLSLIKHPCSTMSGKAIHILAKQGRF 60  
DB 147 MLVRADGGVGVQIIGESTIDVAVGCPDTAKMLKL-PPY-----GGPIREKADAKNGNPH 199  
QY 61 HFIDIKPPLHAKNCDPFTGLQHVTDKIMKKKEKEGIEKQIILSSADIATVQHTAC 120  
DB 200 AYELPRPMQH-KGLDPFSFGKTAIHNIKDTPNAGSDP-----ATRADIASFEYAVVD 253  
QY 121 HLVRTRRAILFCQKRDILLPQNNAVLVASGGVSNFYIRALIELTNAQTCLCPPL 180  
DB 254 TLVAKCTKALQMTGIRKQ-----LVVAGGVSNQMRLRTLTETLRQIDASVYAPTEL 305  
QY 181 CTDNGIMIAMNGIRLRAG 199

DB 306 CTDNGAMIAYAGFCRLSCG 324

Search completed: February 16, 2005, 13:05:57  
Job time : 49.0598 secs

GenCore version 5.1.6 -  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:43:41 ; Search time 14.288 Seconds  
(without alignments)  
1248.675 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVQGVSDPFLIKSLDI.....DISKEVGRASIKVQLKMEI 239

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/6D.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1240	100.0	267	US-10-067-443-22	Sequence 22, Appl
2	1240	100.0	414	US-10-067-443-2	Sequence 2, Appl
3	1213.5	97.9	439	US-10-067-443-19	Sequence 19, Appl
4	326	26.3	463	US-10-067-443-3	Sequence 3, Appl
5	247	19.9	350	US-09-540-236-2726	Sequence 2726, Ap
6	231	18.6	401	US-09-252-991A-17372	Sequence 17372, A
7	212	17.1	357	US-09-543-681A-6513	Sequence 6513, Ap
8	209.5	16.9	342	US-08-087-797-3	Sequence 3, Appl
9	205	16.5	343	US-09-489-039A-9221	Sequence 9221, Ap
10	201.5	16.2	421	US-10-067-443-4	Sequence 4, Appl
11	201.5	16.2	421	US-10-067-443-28	Sequence 28, Appl
12	200	16.1	325	US-08-087-797-2	Sequence 2, Appl
13	197	15.9	363	US-09-107-532A-6609	Sequence 6609, Ap
14	193.5	15.6	327	US-10-067-443-5	Sequence 5, Appl
15	190	15.3	336	US-08-987-121A-4	Sequence 4, Appl
16	189	15.2	273	US-09-710-279-728	Sequence 728, App
17	187	15.1	344	US-09-602-777A-148	Sequence 148, App
18	186	15.0	335	US-08-961-083-52	Sequence 52, Appl
19	186	15.0	335	US-09-536-784-52	Sequence 52, Appl
20	186	15.0	368	US-09-134-001C-3909	Sequence 3909, Ap
21	184	14.8	336	US-09-066-512-2	Sequence 2, Appl
22	182	14.7	336	US-09-583-110-4857	Sequence 4857, Ap
23	181.5	14.6	343	US-09-328-352-4387	Sequence 4387, Ap
24	177	14.3	326	US-05-107-433-4221	Sequence 4221, Ap
25	174.5	14.1	344	US-09-198-452A-213	Sequence 213, App
26	174.5	14.1	360	US-09-438-185A-196	Sequence 196, App
27	174	14.0	341	US-09-149-624-2	Sequence 2, Appl

28	173	14.0	366	US-09-134-000C-4956	Sequence 4956, Ap
29	148.5	12.0	340	US-10-067-443-6	Sequence 6, Appl
30	117	9.4	292	US-09-724-623-81	Sequence 81, Appl
31	94.5	7.6	480	US-09-583-110-5050	Sequence 5050, Ap
32	94.5	7.6	481	US-09-107-433-3197	Sequence 3197, Ap
33	89.5	7.2	214	US-09-328-352-4609	Sequence 4609, Ap
34	83	6.7	42	US-09-902-540-13841	Sequence 13841, A
35	81.5	6.6	1087	US-08-264-002-5	Sequence 5, Appl
36	81	6.5	1072	US-09-902-540-15572	Sequence 15572, A
37	79	6.4	445	US-08-083-945C-2	Sequence 2, Appl
38	79	6.4	445	US-08-083-945C-7	Sequence 7, Appl
39	79	6.4	578	US-09-252-991A-31318	Sequence 31318, A
40	79	6.4	1137	US-09-538-092-968	Sequence 968, App
41	78	6.3	334	US-08-530-950-4	Sequence 4, Appl
42	78	6.3	334	US-08-576-240-2	Sequence 2, Appl
43	78	6.3	334	US-08-888-429A-4	Sequence 4, Appl
44	78	6.3	334	US-09-149-879-4	Sequence 4, Appl
45	78	6.3	334	US-09-057-009-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-10-067-443-22
; Sequence 22, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 22
; LENGTH: 267
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-22

Query Match      100.0%; Score 1240; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 5.6e+145;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LIALVQGVSDPFLIKSLDIAPQDMIDKVARSLIKPECGTMSGKAIIEHLAKQGNRP 60
      |||
DB      29 LIALVQGVSDPFLIKSLDIAPQDMIDKVARSLIKPECGTMSGKAIIEHLAKQGNRF 88
      |||

QY      61 HEDIKPPLHANKCPSPFGIOHVTDKITMKKEKEGKIGQILSSADIAITVQHTMAC 120
      |||
DB      89 HEDIKPPLHANKCPSPFGIOHVTDKITMKKEKEGKIGQILSSADIAITVQHTMAC 148
      |||

QY      121 HLKERTHRAILFCCKRDLLPNNNAVTVASGVAANFYIRALIEILTNATQCTLLCPPL 180
      |||
DB      149 HLKERTHRAILFCCKRDLLPNNNAVTVASGVAANFYIRALIEILTNATQCTLLCPPL 208
      |||

QY      181 CTDNGIMIAMNGIEIRLAGIILHDIIGIRYEPKCELVADISKEVGRASIKVQLKMEI 239
      |||
DB      209 CTDNGIMIAMNGIEIRLAGIILHDIIGIRYEPKCELVADISKEVGRASIKVQLKMEI 267
      |||

RESULT 2
US-10-067-443-2
; Sequence 2, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
```

1 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
2 FILE REFERENCE: D0073 NP  
3 CURRENT APPLICATION NUMBER: US/10/067,443  
4 PRIOR FILING DATE: 2002-02-05  
5 PRIOR APPLICATION NUMBER: US 60/266,518  
6 PRIOR FILING DATE: 2001-02-05  
7 PRIOR APPLICATION NUMBER: US 60/282,814  
8 PRIOR FILING DATE: 2001-04-10  
9 NUMBER OF SEQ ID NOS: 71  
10 SOFTWARE: PatentIn version 3.0  
11 SEQ ID NO 2  
12 LENGTH: 414  
13 TYPE: PRT  
14 ORGANISM: Homo sapiens  
15 US-10-067-443-2

Query Match 100.0%; Score 1240; DB 4; Length 414;  
Best Local Similarity 100.0%; Pred. No. 1.1e-144;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKPECSGTMGGKAIETHLAKQGNRF 60  
DB 176 LLAIVGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKPECSGTMGGKAIETHLAKQGNRF 235  
QY 61 HFDIKPPLHAKNCDPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 120  
DB 236 HFDIKPPLHAKNCDPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 295  
QY 121 HLVKRTHRALIFCQKRDLLPQNNAVLVASGVASNFYIRALEILTNAQCTLLCPPL 180  
DB 296 HLVKRTHRALIFCQKRDLLPQNNAVLVASGVASNFYIRALEILTNAQCTLLCPPL 355  
QY 181 CTDNGIMIAMNGIERLRAGILGILHDIIEGIRYEPKCPPLGVDSIKVYQPKMEI 239  
DB 356 CTDNGIMIAMNGIERLRAGILGILHDIIEGIRYEPKCPPLGVDSIKVYQPKMEI 414

RESULT 3  
US-10-067-443-19  
1 Sequence 19, Application US/10067443  
2 Patent No. 6642041  
3 GENERAL INFORMATION:

4 APPLICANT: Bristol-Myers Squibb Company  
5 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
6 FILE REFERENCE: D0073 NP  
7 CURRENT APPLICATION NUMBER: US/10/067,443  
8 PRIOR FILING DATE: 2002-02-05  
9 PRIOR APPLICATION NUMBER: US 60/266,518  
10 PRIOR FILING DATE: 2001-02-05  
11 PRIOR APPLICATION NUMBER: US 60/282,814  
12 PRIOR FILING DATE: 2001-04-10  
13 NUMBER OF SEQ ID NOS: 71  
14 SOFTWARE: PatentIn version 3.0  
15 SEQ ID NO 19  
16 LENGTH: 439  
17 TYPE: PRT  
18 ORGANISM: homo sapiens  
19 US-10-067-443-19

Query Match 97.9%; Score 1213.5; DB 4; Length 439;  
Best Local Similarity 90.2%; Pred. No. 2.3e-141;  
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 LLAIVGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKPECSGTMGGKAIETHLAKQGNRF 60  
DB 176 LLAIVGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKPECSGTMGGKAIETHLAKQGNRF 235  
QY 61 HFDIKPPLHAKNCDPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 120  
DB 236 HFDIKPPLHAKNCDPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 295

QY 99 ---EKQILSSADIAATVQHTMACHLVYRTHRALIFCQKRDLLPQNNAVLVASGVASN 155  
DB 296 CYEKGQILSSADIAATVQHTMACHLVYRTHRALIFCQKRDLLPQNNAVLVASGVASN 355  
QY 156 FYIRALEILTNAQCTLLCPPLCTDNGIMIAMNGIERLRAGILGILHDIIEGIRYEPK 215  
DB 356 FYIRALEILTNAQCTLLCPPLCTDNGIMIAMNGIERLRAGILGILHDIIEGIRYEPK 415  
QY 216 PLGVDSIKVYQPKMEI 239  
DB 416 PLGVDSIKVYQPKMEI 439

RESULT 4  
US-10-067-443-3  
1 Sequence 3, Application US/10067443  
2 Patent No. 6642041  
3 GENERAL INFORMATION:

4 APPLICANT: Bristol-Myers Squibb Company  
5 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
6 FILE REFERENCE: D0073 NP  
7 CURRENT APPLICATION NUMBER: US/10/067,443  
8 PRIOR FILING DATE: 2002-02-05  
9 PRIOR APPLICATION NUMBER: US 60/266,518  
10 PRIOR FILING DATE: 2001-02-05  
11 PRIOR APPLICATION NUMBER: US 60/282,814  
12 PRIOR FILING DATE: 2001-04-10  
13 NUMBER OF SEQ ID NOS: 71  
14 SOFTWARE: PatentIn version 3.0  
15 SEQ ID NO 3  
16 LENGTH: 463  
17 TYPE: PRT  
18 ORGANISM: Arabidopsis thaliana  
19 US-10-067-443-3

Query Match 26.3%; Score 326; DB 4; Length 463;  
Best Local Similarity 33.6%; Pred. No. 2.7e-31;  
Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;

QY 1 LLAIVGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKPECSGTMGGKAIETHLAKQGNRF 60  
DB 221 LLAIVGVSDPFLILGKSLDIAFGDMLDKVARRLSLIKPECSGTMGGKAIETHLAKQGNRF 275  
QY 61 HFDIKPPLHAKNCDPFTGLQHTVTDKTIIMKKEKEGIEKQILSSADIAATVQHTMAC 120  
DB 276 SVKFNVPKYNKDCNPSYAGLQVRLAIEAKR-----IRNRADIAASFQVAVL 325  
QY 121 HLVKRTHRALIFCQKRDLLPQNNAVLVASGVASNFYIRALEILTNAQCTLLCPPL 180  
DB 326 HLBKCKERAIIDWALR---LEPSIKHNVISGVASNRYRLRLNNIVENKRLKLVCPPL 382  
QY 181 CTDNGIMIAMNGIERLRAGILGILHDIIEGIRYEPKCPPLGVDSIKVYQPKMEI 224  
DB 383 CTDNGIMIAMNGIERLRAGILGILHDIIEGIRYEPKCPPLGVDSIKVYQPKMEI 432  
QY 225 VGEA 228  
DB 433 RSEA 436

RESULT 5  
US-09-540-236-2726  
1 Sequence 2726, Application US/09540236  
2 Patent No. 6673910  
3 GENERAL INFORMATION:

4 APPLICANT: Gary L. Breton et al.  
5 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARI  
6 FILE REFERENCE: 2709.2005-001  
7 CURRENT APPLICATION NUMBER: US/09/540,236  
8 CURRENT FILING DATE: 2000-04-04  
9 NUMBER OF SEQ ID NOS: 3840



SEQ ID NO 2726  
LENGTH: 350  
TYPE: PRT  
ORGANISM: M. catarrhalis  
US-09-540-236-2726

Query Match 19.9%; Score 247; DB 4; Length 350;  
Best Local Similarity 31.5%; Pred. No. 1,1e-21;  
Matches 70; Conservative 34; Mismatches 80; Indels 38; Gaps 6;

QY 1 LALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHEPCSTMSGKAIIEHLAKGNRF 60  
DB 149 MLVADGVGVQIIGESIDDAVGEFCFDTAKLKL-PYP-----GCPNIEKTLAKGNPH 201  
QY 61 HFDIKPPLHAKNCDFSTGQHTVDKTIIMKKEKEGIEKQIILSSADIAATVQHTMAC 120  
DB 202 AYELPRPQOR-KGLDFSPSGMKTAIHNLIKOTPVNAQSP-----ATRADIASFEYAVVD 255  
QY 121 HLVRTRAILFCQKRDILPONNAVLVAGVSNFYIRALEILTNAQTCLCPPL 180  
DB 256 TLVVKCKKALQMTGIRQ-----LVVAGVSNQTLRILTELRQIDASVYVAPREL 307  
QY 181 CTDNGIMIAMNGIERLRAG-----IGILHD 205  
DB 308 CTDNGAMIAVAGFCRLSGQSDILA VRCP RWDMTLGI EYD 349

RESULT 6  
US-09-252-991A-17372  
Sequence 17372, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17372  
LENGTH: 401  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17372

Query Match 18.6%; Score 231; DB 4; Length 401;  
Best Local Similarity 33.8%; Pred. No. 1.3e-19;  
Matches 73; Conservative 34; Mismatches 87; Indels 22; Gaps 8;

QY 2 LALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHEPCSTMSGKAIIEHLAKGNRF 61  
DB 202 LVRVDGIGRYQLDGSVDDAAGAPDKAKLIGL-GYP-----GCPRIARLAEGTGR 254  
QY 62 FDIKPLHAKNCDFSTGQHTVDKTIIMKKEKEGIEKQIILSSADIAATVQHTMAC 121  
DB 255 FVFP RPTDRGLDFSGKTFTFLN-TWRCVEAGDSDSEQ---TRCIALAFQAVVET 310  
QY 122 LVKTRTRAILFCQKRDILPONNAVLVAGVSNFYIRALEILTNAQTCLCPPL 181  
DB 311 LLIKCRBAL---KQGTG---KN---LVLAGVSNQTLRILTELRQIDASVYVAPREL 362  
QY 182 TDNGIMIAMNGIERLRAGLIGILHDIEGIRYEPKPL 217  
DB 363 TDNGAMIAVAGFCRLSGQSDILA VRCP RWDMTLGI EYD 395

RESULT 7  
US-09-543-681A-6513  
Sequence 6513, Application US/09543681A

Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1998-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 6513  
LENGTH: 357  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-6513

Query Match 17.1%; Score 212; DB 4; Length 357;  
Best Local Similarity 30.3%; Pred. No. 2.4e-17;  
Matches 66; Conservative 32; Mismatches 90; Indels 30; Gaps 7;

QY 2 LALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHEPCSTMSGKAIIEHLAKGNRF 61  
DB 159 LSVTGIGBYTLLGSIDDAAGEAPDKAKLIGL-DYP-----GCPVLSKMAQGVETR 211  
QY 62 FDIKPLHAKNCDFSTGQHTVDKTIIMKKEKEGIEKQIILSSADIAATVQHTMAC 121  
DB 212 FVFP RPTDRGLDFSGKTFTFLN-TWRCVEAGDSDSEQ---TRCIALAFQAVVET 263  
QY 122 LVKTRTRAILFCQKRDILPONNAVLVAGVSNFYIRALEILTNAQTCLCPPL 180  
DB 264 LAIKRRA-----LEQTFKRLVAGVSNFYIRALEILTNAQTCLCPPL 314  
QY 181 CTDNGIMIAMNGIERLRAGL-GILHDIEGIRYEPKPL 217  
DB 315 CTDNGAMIAVAGFCRLSGQSDILA VRCP RWDMTLGI EYD 349

RESULT 8

US-08-087-797-3  
Sequence 3, Application US/08087797  
Patent No. 5543312  
GENERAL INFORMATION:

APPLICANT: Mellors, Alan  
APPLICANT: Lo, Reggie Y.C.  
APPLICANT: Abdullah, Khalid M.  
TITLE OF INVENTION: Pasteurella Haemolytica  
TITLE OF INVENTION: Glycoprotease  
TITLE OF INVENTION: Gene and the Purified Enzyme  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Bell, Seltzer, Park & Gibson, P.A.  
STREET: 1211 East Morehead Street,  
CITY: Charlotte  
STATE: No. 5543312th Carolina  
COUNTRY: United States  
ZIP: 28234

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,797  
FILING DATE: 14-JUL-1993  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Layton, Jr., Samuel G.  
REGISTRATION NUMBER: 22907  
REFERENCE/DOCKET NUMBER: 3374-80  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 704 377 1561  
TELEFAX: 704 334 2014

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULAR TYPE: peptide  
US-08-087-797-3

Query Match 16.9%; Score 209.5; DB 1; Length 342;  
Best Local Similarity 30.9%; Pred. No. 4,7e-17;  
Matches 67; Conservative 30; Mismatches 97; Indels 23; Gaps 7;

2 LALVQGVSDPFLLGKSLDIPGDMDKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRH 61  
142 LISTVTGIGQYELLESIDDAAGEAFDKTAKLGL-DYP-----GGPLISKPAAGSTAGR 194  
62 FDIKPLHAKNCDFSFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIATVQHTMACH 121  
195 FVFPFRPMTDRPGLDPSFGSLKTPAANTIRDNXXXXGXTHDQ---TRADIARAFEDAVDT 251  
QY 122 LVKRTTRAILFCQKQDILPQNN-AVLVASGVASNFYIRRALEILTNAQCTLLCPPEPL 180  
DB 252 LMICKKRA-----LDQTFKFLVMAGVGSANRTIRAKIAEMKKKRGSEVFYARPEF 302  
QY 181 CTONGIMAMNGIERLRAGLGIHIDIEGIRYEPKCL 217  
DB 303 CTDNGAMIAVAGWVRLQTA---KAEIGVTVRPRWPL 336

RESULT 9  
US-09-489-039A-9221

Sequence 9221, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709,2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 9221  
LENGTH: 343  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9221

Query Match 16.5%; Score 205; DB 4; Length 343;  
Best Local Similarity 30.3%; Pred. No. 1.7e-16;  
Matches 66; Conservative 29; Mismatches 93; Indels 30; Gaps 7;

2 LALVQGVSDPFLLGKSLDIPGDMDKVARRLSLIKHEPCSTMSGKAIIEHLAKGNRH 61  
148 LISTVTGIGQYELLESIDDAAGEAFDKTAKLGL-DYP-----GGPLISKPAAGSTAGR 200  
QY 62 FDIKPLHAKNCDFSFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIATVQHTMACH 121  
DB 201 FVFPFRPMTDRPGLDPSFGSLKTPAANTIRSNDD-----QTRADIARAFEDAVDT 252  
QY 122 LVKRTTRAILFCQKQDILPQNN-AVLVASGVASNFYIRRALEILTNAQCTLLCPPEPL 180  
DB 253 LMICKKRA-----LEQTFKFLVMAGVGSANRTIRAKIAEMKKKRGSEVFYARPEF 303  
QY 181 CTONGIMAMNGIERLRAGLGIHIDIEGIRYEPKCL 217  
DB 304 CTDNGAMIAVAGWVRLQTA---KAEIGVTVRPRWPL 337

RESULT 10  
US-10-067-443-4

Sequence 4, Application US/10067443  
Patent No. 6642041  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
PRIOR FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 421  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-067-443-4

Query Match 16.2%; Score 201.5; DB 4; Length 421;  
Best Local Similarity 27.5%; Pred. No. 6.3e-16;  
Matches 69; Conservative 48; Mismatches 99; Indels 35; Gaps 7;

1 LIALVQGVSDPFLLGKSLDIPGDMDKVARRLSLIKHEPCSTMSGKAIIEHLAKGN-- 58  
163 LITVADVEKFLYQSGSPECIDKVARQGD-LGSEFGQIHGAIVELIASPASD 221  
QY 59 -RFHFDIKPLHAKNCDFSFTGLQHTVDKIIIMKKEKEGIEKGQILSSAADIATVQHT 117  
DB 222 GHLRYPILPVPKXNMNDQIKGSYANLELRKNSFISID-----IPFCASLQNT 274  
QY 118 MACHLVKRTTRAILFCQKQDILPQNNAVLVASGVASNFYIRRALEILTNAQCTLLCP 177  
DB 275 VARHISSKLHIFPESISEQEKLPKQ---LVIGGVANQYIRGALSLSAAHVITIKYL 331  
QY 178 PRLCTDNGIMAMNGIERLRAGLGIHIDIEGIRYEP-----KCPLGVDISKEVG 226  
DB 332 LSLCTDNAMEIAVSGL-----LMLVNRSEALIMWRNDIPDTIYAHARSDIGDASSET 384  
QY 227 EASIKVPOLKM 237  
DB 385 ---IDTPRRKL 392

RESULT 11

US-10-067-443-28  
Sequence 28, Application US/10067443

Patent No. 6642041  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
PRIOR FILING DATE: 2002-02-05  
PRIOR APPLICATION NUMBER: US 60/266,518  
PRIOR FILING DATE: 2001-02-05  
PRIOR APPLICATION NUMBER: US 60/282,814  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 28  
LENGTH: 421  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-067-443-28

Query Match 16.2%; Score 201.5; DB 4; Length 421;  
Best Local Similarity 27.5%; Pred. No. 6.3e-16;  
Matches 69; Conservative 48; Mismatches 99; Indels 35; Gaps 7;

QY 1 LIALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSWSGKAI EHLAKQGN-- 58  
Db 163 LISVAEVEKFKLYGVSGSGSPGECIDKVARQLDGL-GSEFDGIVGAIVELASRASD 221  
QY 59 -RHFEDIKPLPHANCDPFTGLQHTDKIMKEKEBGEIKGQILSSADIAATVQHT 117  
Db 222 GHLRYPFLPRVPKANNFDDIKGSYLNILRLRKNSETSID-----IPDFCASLQNT 274  
QY 118 MACLVKTRRALIFCQKORDLPQNNAVLVASGVASNFYIRALLETITNATQCTLLCP 177  
Db 275 VARIHSSKLIHFPFSLSEQEDLPKQ--LVIGGVANQYLFGLISKLSAANNVTITKXL 331  
QY 178 PRLCTDNGIMIANNGIERLRAGLILHDIGIRYEP-----KCPGLGVDSKEVG 226  
Db 332 LSLCTDAEMIAVSGL-----LMLVNRSEALIMWRPNIPDTYAHASDICTDASSER- 384  
QY 227 EASIKVPOLKM 237  
Db 385 ---IDTPRRKL 392  
RESULT 12  
US-08-087-797-2  
; Sequence 2, Application US/08087797  
; Patent No. 5543312  
; GENERAL INFORMATION:  
; APPLICANT: Mellore, Alan  
; APPLICANT: Lo, Reggie Y.C.  
; APPLICANT: Abdullah, Khalid M.  
; TITLE OF INVENTION: Pasteurella Haemolytica  
; TITLE OF INVENTION: Glycoprotease  
; TITLE OF INVENTION: Gene and the Purified Enzyme  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.  
; STREET: 1211 East Morehead Street,  
; CITY: Charlotte  
; STATE: No. 5543312th Carolina  
; COUNTRY: United States  
; ZIP: 28234  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PC-DOS/MS-DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/087,797  
; FILING DATE: 14-JUL-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Layton, Jr., Samuel G.  
; REGISTRATION NUMBER: 22807  
; REFERENCE/DOCKET NUMBER: 3374-80  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 704 377 1561  
; TELEFAX: 704 334 2014  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 325 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-087-797-2  
Query Match 16.1%; Score 200; DB 1; Length 325;  
Best Local Similarity 31.6%; Pred. No. 6.5e-16;  
Matches 62; Conservative 25; Mismatches 91; Indels 18; Gaps 5;  
QY 2 LALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSWSGKAI EHLAKQGNF 61  
Db 142 LVVVDGCGYBILGESIDDAAGEAFDKTKLGL-DY-----AGVMSKLAEGSTNR 194  
QY 62 FDIKPLPHANCDPFTGLQHTDKIMKEKEBGEIKGQILSSADIAATVQHTMAC 121

Db 195 FEPFRMTDRPGIDFSGLKTFEANTIKANINENDELDEQ--TKCDIAHAFQAV-- 248  
QY 122 LKRTTRALIFCQKORDLPQNNAVLVASGVASNFYIRALLETITNATQCTLLCP 181  
Db 249 ---VDTLIKCK-RALBETGYKRLVMAGVSNKQLRADLAEWMKKLKEGVYFPPQFC 303  
QY 182 TDNGIMIANNGIERLR 197  
Db 304 TDNGAMIAATGFLRLK 319  
RESULT 13  
US-09-107-532A-6609  
; Sequence 6609, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 6609:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 363 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (8) LOCATION 1...363  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6609:  
US-09-107-532A-6609  
Query Match 15.9%; Score 197; DB 4; Length 363;  
Best Local Similarity 32.6%; Pred. No. 1.8e-15;  
Matches 63; Conservative 34; Mismatches 64; Indels 32; Gaps 9;  
QY 2 LALVGVSDFLILGKSLDIAPGMDLKVARRSLIKHPECSWSGKAI EHLAKQGNF 60  
Db 169 LVVVDGDSYBILIGETBDDAAGEAYDKGRVIGL-----SYPSGKIDQLAHQCKDNY 221  
QY 61 HFDIKPLPHANCDPFTGLQHTDKIMKEKEBGEIKGQILSSADIAATVQHTMA 119

Db 222 HF--PRAMIHEDNDYDFSGSLKSAFINLVHNAOGRGDLJDN-----DLAASFOASVT 272  
QY 120 CHVKTGTRATLFCQKQDRLPONNAV--LVASGVASNFYIRALBITLNA--TQCTLLC 175  
Db 273 DVLTKTLRA---C-----QNPVKQLVAVGVAAVNAOGLBGLQALMSAKLPEVELVT 322  
QY 176 PPPRLCTDNGIMI 188  
Db 323 PPLRLCGDNAMMI 335

RESULT 14  
US-10-067-443-5  
; Sequence 5, Application US/10067443  
; Patent No. 6642041

; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 327  
; TYPE: PRT  
; ORGANISM: Thermotoga maritima  
US-10-067-443-5

Query Match 15.6%; Score 193.5; DB 4; Length 327;  
Best Local Similarity 28.9%; Pred. No. 4.2e-15;  
Matches 55; Conservative 38; Mismatches 68; Indels 29; Gaps 6;

QY 13 LIGSLDIAPGMDLKVARRSLIKHPECSTMSGKALIEHLAKQGNRRHFIDIKPPLHNAK 72  
Db 151 VLGETLDSAGEAPDKVARLIGL-GYP-----GGPVIDRAVKGDPEKYSFPRPMLDD 203  
QY 73 NCDSFPTGLQHTDKIIMKKEKEGIEKQILSSAADIATVQHTMACHLVKTRATILF 132  
Db 204 SYNSFAGLK--TSVLYFLQREKGYK-----VEDVAASFQKAVDILVEKTR---- 249  
QY 133 CKQDRLPONNAV--LVASGVASNFYIRALBITLNACTCLCPPLCTDNGIMIAM 190  
Db 250 -----LARNIGIRKIAVGVAAVNAOGLBGLQALMSAKLPEVELVT 322  
QY 191 NGIERLRAQL 200  
Db 304 AGYEKAKRGM 313

RESULT 15

US-08-987-121A-4  
; Sequence 4, Application US/08987121A  
; GENERAL INFORMATION:

; APPLICANT: Hoskins, Jo Ann  
; APPLICANT: Tang, Joseph Chioy-Chung  
; APPLICANT: Treadway, Patil Jean  
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: U.S.  
; ZIP: 46285  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/987,121A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Webster, Thomas D.  
; REGISTRATION NUMBER: 39,872  
; REFERENCE/DOCKET NUMBER: X-  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3334  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 336 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-987-121A-4

Query Match 15.3%; Score 190; DB 3; Length 336;  
Best Local Similarity 31.1%; Pred. No. 1.2e-14;  
Matches 61; Conservative 32; Mismatches 73; Indels 30; Gaps 8;

QY 2 LALVGVSDPFLIGSLDIAPGMDLKVARRSLIKHPECSTMSGKALIEHLAKQGNRRH 61  
Db 144 LVYVSBAGDYKIVGETRDVAGEAYDKGVVNGL-----TYPAGREIDELAHQCHDI- 195  
QY 62 FDIKPELHNAKNCDSFPTGLQHTDKIIMKKEKEGIEKQILSSAADIATVQHTMACH 121  
Db 196 YDPPRAMIKEDYLBRSFSGLSKSAFINHNAE-----QNGESLST-EDLCASQAANVMDI 249  
QY 122 LVKTRTRATLFCQKQDRLPONNAV--LVASGVASNFYIRALBITLNACTCLCPPLR 181  
Db 250 LMAKTKKAL-----EKYEVK--TLVAVAGVAAVNAOGLBGLQALMSAKLPEVELVT 299

QY 182 TDNGIMIA-----WN 191  
Db 300 GDNAGMIAYASVSEWN 315

Search completed: February 16, 2005, 13:10:26  
Job time : 15.288 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 13:08:47 ; Search time 138.464 Seconds  
(without alignments)  
563.995 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVQGVSDPFLILGKSLDI.....DISKVEGASIKVPLKMEI 239

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: Published Applications\_AA:\*  
2: /cgn2\_6/ptcodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptcodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptcodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptcodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptcodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptcodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptcodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptcodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptcodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptcodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptcodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptcodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptcodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptcodata/2/pubpaa/US10D\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptcodata/2/pubpaa/US10E\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptcodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptcodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptcodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptcodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1240	100.0	267 14	US-10-067-443-22 Sequence 22, Appl
2	1240	100.0	267 15	US-10-649-273-22 Sequence 22, Appl
3	1240	100.0	267 15	US-10-651-722-22 Sequence 22, Appl
4	1240	100.0	414 14	US-10-067-443-22 Sequence 2, Appl
5	1240	100.0	414 15	US-10-649-273-22 Sequence 2, Appl
6	1240	100.0	414 15	US-10-651-722-22 Sequence 2, Appl
7	1213.5	97.9	439 14	US-10-067-443-19 Sequence 19, Appl
8	1213.5	97.9	439 15	US-10-649-273-19 Sequence 19, Appl
9	1213.5	97.9	439 15	US-10-651-722-19 Sequence 19, Appl
10	1202	97.0	414 14	US-10-012-140-5 Sequence 5, Appl
11	960	77.4	364 14	US-10-094-749-2039 Sequence 2039, Ap
12	337	27.2	445 15	US-10-424-599-209259 Sequence 209259, Ap
13	326	26.3	463 14	US-10-067-443-3 Sequence 3, Appl

14	326	26.3	463 15	US-10-649-273-3 Sequence 3, Appl
15	326	26.3	463 15	US-10-651-722-3 Sequence 3, Appl
16	270	21.8	444 15	US-10-437-963-113732 Sequence 113732, A
17	257	20.7	382 15	US-10-282-122A-50858 Sequence 50858, A
18	254	20.5	333 14	US-10-012-140-25 Sequence 25, Appl
19	248.5	20.0	338 15	US-10-282-122A-63156 Sequence 63156, A
20	247	19.9	312 15	US-10-282-122A-44499 Sequence 44499, A
21	241	19.4	343 15	US-10-282-122A-67227 Sequence 67227, A
22	239	18.6	251 9	US-10-081-051-9 Sequence 9, Appl
23	231	18.6	341 15	US-09-815-242-11798 Sequence 11798, A
24	231	18.6	341 15	US-10-282-122A-66200 Sequence 66200, A
25	228.5	18.4	347 14	US-10-012-140-24 Sequence 24, Appl
26	224	18.1	342 9	US-09-815-242-11043 Sequence 11043, A
27	224	18.1	342 15	US-10-282-122A-58204 Sequence 58204, A
28	213	17.2	337 15	US-09-815-242-13780 Sequence 13780, A
29	213	17.2	337 15	US-10-282-122A-75485 Sequence 75485, A
30	212	17.1	340 15	US-10-282-122A-68438 Sequence 68438, A
31	211	17.0	41 14	US-10-012-140-26 Sequence 26, Appl
32	210	16.9	337 15	US-10-282-122A-78161 Sequence 78161, A
33	208	16.8	335 15	US-10-282-122A-55404 Sequence 55404, A
34	207.5	16.7	341 15	US-10-282-122A-67993 Sequence 67993, A
35	207	16.7	441 16	US-10-437-963-111620 Sequence 111620, A
36	206	16.6	337 9	US-09-815-242-10304 Sequence 10304, A
37	206	16.6	337 15	US-10-282-122A-56695 Sequence 56695, A
38	206	16.6	337 15	US-10-425-114-67198 Sequence 67198, A
39	201.5	16.2	421 14	US-10-067-443-4 Sequence 4, Appl
40	201.5	16.2	421 14	US-10-649-273-4 Sequence 28, Appl
41	201.5	16.2	421 15	US-10-649-273-4 Sequence 28, Appl
42	201.5	16.2	421 15	US-10-649-273-28 Sequence 28, Appl
43	201.5	16.2	421 15	US-10-651-722-4 Sequence 4, Appl
44	201.5	16.2	421 15	US-10-651-722-28 Sequence 28, Appl
45	199.5	16.1	350 15	US-10-282-122A-53665 Sequence 53665, A

ALIGNMENTS

RESULT 1  
US-10-067-443-22  
; Sequence 22, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-067-443-22  
Query Match 100.0%; Score 1240; DB 14; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.8e-127; Mismatches 0; Indels 0; Gaps 0;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LIALVQGVSDPFLILGKSLDIAPGMDLVKVARLSLIKHEGCTMSGKAIIEHLAKQGRF 60  
DB 29 LIALVQGVSDPFLILGKSLDIAPGMDLVKVARLSLIKHEGCTMSGKAIIEHLAKQGRF 88  
QY 61 HFDIPEPLHAKNGCPSPFQIYVTDKIKKKEGEGIEKQILSSAADIATVQHTMAC 120  
DB 89 HFDIPEPLHAKNGCPSPFQIYVTDKIKKKEGEGIEKQILSSAADIATVQHTMAC 148  
QY 121 HLVKTRHALLFCRQDRLIPONNAVIVASGVASNFYIRRALEILTNATQITLLCPPL 180

Db 149 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 208  
QY 181 CTDNGIMTAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEASIKVPOLKMEI 239  
Db 209 CTDNGIMTAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEASIKVPOLKMEI 267

## RESULT 2

US-10-649-273-22  
; Sequence 22, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT APPLICATION NUMBER: US/10/649,273  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 22  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-649-273-22

Query Match 100.0%; Score 1240; DB 15; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.8e-127;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMGGKAI EHLAKGNRF 60  
Db 29 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMGGKAI EHLAKGNRF 88  
QY 61 HFDIKPPLHAHAKNCDPFTGLQHVTDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120  
Db 89 HFDIKPPLHAHAKNCDPFTGLQHVTDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 148  
QY 121 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 180  
Db 149 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 208  
QY 181 CTDNGIMTAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEASIKVPOLKMEI 239  
Db 209 CTDNGIMTAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEASIKVPOLKMEI 267

## RESULT 3

US-10-651-722-22  
; Sequence 22, Application US/10651722  
; Publication No. US20040048302A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1  
; FILE REFERENCE: D0073 DIV  
; CURRENT APPLICATION NUMBER: US/10/651,722  
; PRIOR FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 22  
; LENGTH: 267

; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-651-722-22

Query Match 100.0%; Score 1240; DB 15; Length 267;  
Best Local Similarity 100.0%; Pred. No. 1.8e-127;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMGGKAI EHLAKGNRF 60  
Db 29 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMGGKAI EHLAKGNRF 88  
QY 61 HFDIKPPLHAHAKNCDPFTGLQHVTDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120  
Db 89 HFDIKPPLHAHAKNCDPFTGLQHVTDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 148  
QY 121 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 180  
Db 149 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 208  
QY 181 CTDNGIMTAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEASIKVPOLKMEI 239  
Db 209 CTDNGIMTAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEASIKVPOLKMEI 267

## RESULT 4

US-10-067-443-2  
; Sequence 2, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-067-443-2

Query Match 100.0%; Score 1240; DB 14; Length 414;  
Best Local Similarity 100.0%; Pred. No. 3.4e-127;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMGGKAI EHLAKGNRF 60  
Db 176 LALVQGVSDPFLILGKSLDIAPGMDLKVARRLSLIKHEPCSTMGGKAI EHLAKGNRF 235  
QY 61 HFDIKPPLHAHAKNCDPFTGLQHVTDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120  
Db 236 HFDIKPPLHAHAKNCDPFTGLQHVTDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 295  
QY 121 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 180  
Db 296 HLVKRTHRALIFCKQRDLIPQNNAVLVASGVASNFYIRRALEILTNATQCTLLCPPRL 355  
QY 181 CTDNGIMTAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEASIKVPOLKMEI 239  
Db 356 CTDNGIMTAMNGIERLRAGLIGLHDIEGIRYRKPCLGVDISKVEASIKVPOLKMEI 414

## RESULT 5

US-10-649-273-2  
; Sequence 2, Application US/10649273  
; Publication No. US20040043407A1

```

; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-649-273-2

```

```

Query Match      100.0%; Score 1240; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.4e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 LLAIVGVSDPFLLLGKSLDIAPGMDLKVARRLSLIKHEPCSTMGSKAIEHLAKQGNF 60
Db 176 LLAIVGVSDPFLLLGKSLDIAPGMDLKVARRLSLIKHEPCSTMGSKAIEHLAKQGNF 235
Qy 61 HFDIKPPLHAKNCDPSTGLQHTVTDKTIIMKKEKEGIEKQIILSSADIAATVQHTMAC 120
Db 236 HFDIKPPLHAKNCDPSTGLQHTVTDKTIIMKKEKEGIEKQIILSSADIAATVQHTMAC 295
Qy 121 HLVRTRHAILFCQKQDILLPNNNAVVASGVASNFYIRRALEILTNAQTCTLLCPPL 180
Db 296 HLVRTRHAILFCQKQDILLPNNNAVVASGVASNFYIRRALEILTNAQTCTLLCPPL 355
Qy 181 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKPLGVDISKVEGASIKVPOLKMEI 239
Db 356 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKPLGVDISKVEGASIKVPOLKMEI 414

```

```

RESULT 6
; Sequence 2, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-651-722-2

```

```

Query Match      100.0%; Score 1240; DB 15; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.4e-127;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 LLAIVGVSDPFLLLGKSLDIAPGMDLKVARRLSLIKHEPCSTMGSKAIEHLAKQGNF 60
Db 176 LLAIVGVSDPFLLLGKSLDIAPGMDLKVARRLSLIKHEPCSTMGSKAIEHLAKQGNF 235

```

```

Qy 61 HFDIKPPLHAKNCDPSTGLQHTVTDKTIIMKKEKEGIEKQIILSSADIAATVQHTMAC 120
Db 236 HFDIKPPLHAKNCDPSTGLQHTVTDKTIIMKKEKEGIEKQIILSSADIAATVQHTMAC 295
Qy 121 HLVRTRHAILFCQKQDILLPNNNAVVASGVASNFYIRRALEILTNAQTCTLLCPPL 180
Db 296 HLVRTRHAILFCQKQDILLPNNNAVVASGVASNFYIRRALEILTNAQTCTLLCPPL 355
Qy 181 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKPLGVDISKVEGASIKVPOLKMEI 239
Db 356 CTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPKPLGVDISKVEGASIKVPOLKMEI 414

```

```

RESULT 7
; Sequence 19, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-067-443-19

```

```

Query Match      97.9%; Score 1213.5; DB 14; Length 439;
Best Local Similarity 90.2%; Pred. No. 3e-124; 1; Indels 25; Gaps 1;
Matches 238; Conservative 0; Mismatches 1;

```

```

Qy 1 LLAIVGVSDPFLLLGKSLDIAPGMDLKVARRLSLIKHEPCSTMGSKAIEHLAKQGNF 60
Db 176 LLAIVGVSDPFLLLGKSLDIAPGMDLKVARRLSLIKHEPCSTMGSKAIEHLAKQGNF 235
Qy 61 HFDIKPPLHAKNCDPSTGLQHTVTDKTIIMKKEKEGIEKQIILSSADIAATVQHTMAC 120
Db 236 HFDIKPPLHAKNCDPSTGLQHTVTDKTIIMKKEKEGIEKQIILSSADIAATVQHTMAC 295
Qy 99 ---EKQIILSSADIAATVQHTMACHLVTRTRAILFCQKQDILLPNNNAVVASGVASN 155
Db 296 CRYEKGQIILSSADIAATVQHTMACHLVTRTRAILFCQKQDILLPNNNAVVASGVASN 355
Qy 156 FYIRRALEILTNAQTCTLLCPPLCTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPK 215
Db 356 FYIRRALEILTNAQTCTLLCPPLCTDNGIMIAMNGIERLRAGLGIHDIIEGIRYEPK 415
Qy 216 PLGVDISKVEGASIKVPOLKMEI 239
Db 416 PLGVDISKVEGASIKVPOLKMEI 439

```

```

RESULT 8
; Sequence 19, Application US/10649273
; Publication No. US20040043407A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 CNT
; CURRENT APPLICATION NUMBER: US/10/649,273
; CURRENT FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05

```

```

; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-649-273-19

```

```

Query Match      97.9%; Score 1213.5; DB 15; Length 439;
Best Local Similarity 90.2%; Pred. No. 3e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

```

```

QY 1 LLAIVQGVSDFLILGKSLDIAPGMDLVKVARRLSLIKHEPCSTMSGGKAI EHLAKGNRF 60
DB 176 LLAIVQGVSDFLILGKSLDIAPGMDLVKVARRLSLIKHEPCSTMSGGKAI EHLAKGNRF 235
QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGI----- 98
DB 236 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGI.FLISKVQINIPGLCLKIAHP 295
QY 99 ---EKQGLSSAADIAATVQHTMACHLVKRTTRAILFCQKRDLPONNAVIVASGVASN 155
DB 296 CRYEKGLSSAADIAATVQHTMACHLVKRTTRAILFCQKRDLPONNAVIVASGVASN 355
QY 156 FYIRRALBITNATQCTLLCPPRLCTDNGIMIAMNGIERLRAGILHDIGIRYEPKC 215
DB 356 FYIRRALBITNATQCTLLCPPRLCTDNGIMIAMNGIERLRAGILHDIGIRYEPKC 415
QY 216 PLGVDISKEVGASIKVPOLKMEI 239
DB 416 PLGVDISKEVGASIKVPOLKMEI 439

```

```

RESULT 9
US-10-651-722-19
; Sequence 19, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; PRIOR FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 439
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-651-722-19

```

```

Query Match      97.9%; Score 1213.5; DB 15; Length 439;
Best Local Similarity 90.2%; Pred. No. 3e-124;
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

```

```

QY 99 ---EKQGLSSAADIAATVQHTMACHLVKRTTRAILFCQKRDLPONNAVIVASGVASN 155
DB 296 CRYEKGLSSAADIAATVQHTMACHLVKRTTRAILFCQKRDLPONNAVIVASGVASN 355
QY 156 FYIRRALBITNATQCTLLCPPRLCTDNGIMIAMNGIERLRAGILHDIGIRYEPKC 215
DB 356 FYIRRALBITNATQCTLLCPPRLCTDNGIMIAMNGIERLRAGILHDIGIRYEPKC 415
QY 216 PLGVDISKEVGASIKVPOLKMEI 239
DB 416 PLGVDISKEVGASIKVPOLKMEI 439

```

```

RESULT 10
US-10-012-140-5
; Sequence 5, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leibny, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-140-5

```

```

Query Match      97.0%; Score 1203; DB 14; Length 414;
Best Local Similarity 97.1%; Pred. No. 3.9e-123;
Matches 232; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 LLAIVQGVSDFLILGKSLDIAPGMDLVKVARRLSLIKHEPCSTMSGGKAI EHLAKGNRF 60
DB 176 LLAIVQGVSDFLILGKSLDIAPGMDLVKVARRLSLIKHEPCSTMSGGKAI EHLAKGNRF 235
QY 61 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIKGLSSAADIAATVQHTMAC 120
DB 236 HFDIKPPLHAKNCDFSTGLQHTVDKTIIMKKEKEGIKGLSSAADIAATVQHTMAC 295
QY 121 HLVKTRTRAILFCQKRDLPONNAVIVASGVASN FYIRRALBITNATQCTLLCPPRL 180
DB 296 HLVKTRTRAILFCQKRDLPONNAVIVASGVASN FYIRRALBITNATQCTLLCPPRL 355
QY 181 CTDNGIMIAMNGIERLRAGILHDIGIRYEPKPLGVDISKEVGASIKVPOLKMEI 239
DB 356 CTDNGIMIAMNGIERLRAGILHDIGIRYEPKPLGVDISKEVGASIKVPOLKMEI 414

```

```

RESULT 11
US-10-094-749-2039
; Sequence 2039, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAL, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO

```



APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: -ISONO, YUUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKA, ICHIRO  
APPLICANT: SEKI, NAOHICO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTOKU  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, YASUHIKO  
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
FILE REFERENCE: 084335/0160  
CURRENT APPLICATION NUMBER: US/10/094,749  
CURRENT FILING DATE: 2002-03-12  
PRIORITY FILING DATE: 2002-01-24  
PRIORITY FILING DATE: 2002-01-24  
PRIORITY FILING DATE: 2001-09-14  
PRIORITY FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2039  
LENGTH: 364  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-094-749-2039

Query Match 77.4%; Score 960; DB 15; Length 364;  
Best Local Similarity 98.4%; Pred. No. 1.7e-96;  
Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LLAIVGVSDPFLGKSLDIAPGMDLVARRLSLIKHPECSTMGSGKAIIEHLAQGNRF 60  
DB 176 LLAIVGVSDPFLGKSLDIAPGMDLVARRLSLIKHPECSTMGSGKAIIEHLAQGNRF 235  
QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKTIIMKEKEGIEKGQILSSADIAATVQHTMAC 120  
DB 236 HFDIKPPLHAKNCDPFTGLQHTVDKTIIMKEKEGIEKGQILSSADIAATVQHTMAC 295  
QY 121 HLVRTRRAILFCQKRDILLPNNNAVLVAGGVASNFYIRRALEILTNAQCTLLCPPEPL 180  
DB 296 HLVRTRRAILFCQKRDILLPNNNAVLVAGGVASNFYIRRALEILTNAQCTLLCPPEPL 355  
QY 181 CTDNGIMIA 189  
DB 356 CTDNGIMIA 364

## RESULT 12

US-10-424-599-209259  
Sequence 209259, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(51323)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 209259  
LENGTH: 445  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_3098C.1.pep  
US-10-424-599-209259

Query Match 27.2%; Score 337; DB 15; Length 445;

Best Local Similarity 38.5%; Pred. No. 6.7e-28;  
Matches 85; Conservative 29; Mismatches 83; Indels 24; Gaps 5;

QY 1 LLAIVGVSDPFLGKSLDIAPGMDLVARRLSLIKHPECSTMGSGKAIIEHLAQGNRF 60  
DB 194 LLAIVGVSDPFLGKSLDIAPGMDLVARRLSLIKHPECSTMGSGKAIIEHLAQGNRF 248  
QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKTIIMKEKEGIEKGQILSSADIAATVQ 115  
DB 249 SVKFSIPMQLHNDKNSVAGLTKQVLALESKIDAKIPSSASNGDL-SRADIASFQ 307  
QY 116 HTMAHLVRTRRAILFCQKRDILLPNNNAVLVAGGVASNFYIRRALEILTNAQCTLLC 175  
DB 308 RLAIVHLERECERALQWALKEBPSIRH-LVVGGVASNQVYRARLDWVKKNGQLVC 364  
QY 176 PPRRLCTDNGIMIANNGIERLAGLILDEIRPEKCP 216  
DB 365 PPRRLCTDNGIMIANNGIERLAGLILDEIRPEKCP 395

## RESULT 13

US-10-067-443-3  
Sequence 3, Application US/10067443  
Publication No. US20030082782A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
FILE REFERENCE: D0073 NP  
CURRENT APPLICATION NUMBER: US/10/067,443  
CURRENT FILING DATE: 2002-02-05  
PRIORITY FILING DATE: 2002-02-05  
PRIORITY FILING DATE: 2001-02-05  
PRIORITY FILING DATE: 2001-02-05  
PRIORITY FILING DATE: 2001-04-10  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3  
LENGTH: 463  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-10-067-443-3

Query Match 26.3%; Score 326; DB 14; Length 463;  
Best Local Similarity 33.6%; Pred. No. 1.1e-26;  
Matches 82; Conservative 34; Mismatches 84; Indels 44; Gaps 5;

QY 1 LLAIVGVSDPFLGKSLDIAPGMDLVARRLSLIKHPECSTMGSGKAIIEHLAQGNRF 60  
DB 221 LLAIVGVSDPFLGKSLDIAPGMDLVARRLSLIKHPECSTMGSGKAIIEHLAQGNRF 275  
QY 61 HFDIKPPLHAKNCDPFTGLQHTVDKTIIMKEKEGIEKGQILSSADIAATVQHTMAC 120  
DB 276 SVKFSIPMQLHNDKNSVAGLTKQVLALESKIDAKIPSSASNGDL-SRADIASFQ 325  
QY 121 HLVRTRRAILFCQKRDILLPNNNAVLVAGGVASNFYIRRALEILTNAQCTLLCPPEPL 180  
DB 326 HLEKCEKRAIDWALR-LEPSIKHNVISGVASNFYIRRALEILTNAQCTLLCPPEPL 382  
QY 181 CTDNGIMIANNGIERLAGLILDEIRPEKCP 216  
DB 383 CTDNGIMIANNGIERLAGLILDEIRPEKCP 395  
QY 225 VGEA 228  
DB 433 RSEA 436

## RESULT 14

US-10-649-273-3  
Sequence 3, Application US/10649273  
Publication No. US20040043407A1  
GENERAL INFORMATION:



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:10:32 ; Search time 3803.22 Seconds  
(without alignments)  
3045.000 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVGVDFLLGLKSLDI.....DISKVGESAKIVPQLKMEI 239

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 segs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn21/USPTO.spool/US10649273/runat\_14022005\_114702\_16399/app\_query.fasta\_1.1429  
-DB=GenEmbl -OPMT=faetap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human0.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NORM=exc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US10649273 @CGN\_1\_1 8655 @runat\_14022005\_114702\_16399 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : GenEmbl.\*  
1: gb ba:\*  
2: gb hrg:\*  
3: gb in:\*  
4: gb om:\*  
5: gb ov:\*  
6: gb pat:\*  
7: gb ph:\*  
8: gb pl:\*  
9: gb pr:\*  
10: gb ro:\*  
11: gb sts:\*  
12: gb sy:\*  
13: gb un:\*  
14: gb vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	1240	100.0	1416	6 AR541929 Sequence
2	1240	100.0	1526	6 AR428809 Sequence
3	1240	100.0	1908	9 BC011904 Homo sapi
4	1240	100.0	2197	6 AR428803 Sequence

5	1213.5	97.9	1387	6	AR428808	AR428808 Sequence
6	1213.5	97.9	1387	6	HS4295148	AJ295148 Homo sapi
7	1203	97.0	1245	6	AX664697	AX664697 Sequence
8	1203	97.0	1820	6	AX664695	AX664695 Sequence
9	1086	87.6	1844	10	BC058172	BC058172 Mus muscu
10	1079	87.0	1017	10	BC038910	BC038910 Mus muscu
11	1059	85.4	2208	6	AX713716	AX713716 Sequence
12	1059	85.4	2208	6	AK055441	AK055441 Homo sapi
13	983	79.3	1546	10	BC078974	BC078974 Rattus no
14	944.5	76.2	1109	10	BC051211	BC051211 Mus muscu
15	809	65.2	860	5	BX930694	BX930694 Gallus ga
16	806	65.0	1522	5	BX934991	BX934991 Gallus ga
17	803	64.8	1558	5	BX930963	BX930963 Gallus ga
18	725	58.5	14364	6	AR428807	AR428807 Sequence
19	725	58.5	84115	9	AC013468	AC013468 Homo sapi
20	666	53.7	249601	2	AC114153	AC114153 Rattus no
21	666	53.7	308652	2	AC121478	AC121478 Rattus no
22	646.5	52.1	256751	10	AC122925	AC122925 Mus muscu
23	610	49.2	875	6	CQ721898	CQ721898 Sequence
24	384.5	31.0	1474	3	AK113378	AK113378 Clona int
25	369.5	29.8	121251	5	AL591593	AL591593 Zebrafish
26	356	28.7	20021	10	AF367967	AF367967 Mus muscu
27	356	28.7	179252	10	AF131205	AF131205 Mus muscu
28	345	27.8	117322	5	AL672217	AL672217 Zebrafish
29	332.5	26.8	1576	3	AY051882	AY051882 Drosophila
30	332.5	26.8	1601	6	CQ606432	CQ606432 Sequence
31	332.5	26.8	3656	6	CQ606431	CQ606431 Sequence
32	332.5	26.8	14679	2	AC018262	AC018262 Drosophila
33	332.5	26.8	180263	3	AC010671	AC010671 Drosophila
34	332.5	26.8	207432	3	AE003513	AE003513 Drosophila
35	324	26.1	1443	8	AY024338	AY024338 Arabidops
36	324	26.1	1474	8	AY117283	AY117283 Arabidops
37	324	26.1	1567	8	AY063864	AY063864 Arabidops
38	322	26.0	1577	8	AY084577	AY084577 Arabidops
39	311.5	25.1	1672	8	AK070912	AK070912 Oryza sat
40	305	24.6	110000	2	AP006501	AP006501 Oryza sat
41	262.5	21.2	333800	1	SME551792	Continuation (9 of
42	260	21.0	1718	8	AK099965	AK099965 Oryza sat
43	257	20.7	349354	1	BX640416	BX640416 Bordetella
44	254	20.5	260271	1	AE017258	AE017258 Wolbachia
45	253	20.4	349116	1	AP003003	AP003003 Mesorhizo

ALIGNMENTS

RESULT 1	AR541929	1416 bp	DNA	linear	PAT 08-OCT-2004
LOCUS	AR541929	Sequence 177 from patent US 6743619.			
DEFINITION	AR541929				
ACCESSION	AR541929.1	GI:53934009			
VERSION	AR541929.1				
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 1416)				
AUTHORS	Tang, Y.T., Zhou, P., Goodrich, R., Liu, C., Asundi, V., Ren, F., Zhang, J., Zhao, Q.A., Yang, Y., Xue, A.D., Wehrman, T., Wang, J.-R., Wang, D. and Drmanac, R.T.				
TITLE	Nucleic acids and polypeptides				
JOURNAL	Patent: US 6743619-A 177 01-JUN-2004;				
FEATURES	Location/Qualifiers				
source	1..1416				
ORIGIN	/organism="unknown"				
	/mol_type="genomic DNA"				

Alignment Scores:  
Pred. No.: 2,240,113  
Score: 1240.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
Length: 1416  
Matches: 239  
Conservative: 0  
Mismatch: 0  
Indels: 0

DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x ARS41929 (1-1416)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
 DB 586 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCGCTTCTTGAAAGCTTTGGACATA 645

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgAspLeuSerLeuIleLysHisProGlu 40  
 DB 646 GCACGAGGTGACATGCTTGACAGAGGTGGCAAGAACCTTTCTTAATAAATCAATCCAGG 705

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
 DB 706 TGCTCCACCATGAGTGTGGGAAAGCCATAGAACCTTTGGCCAAACAGGAATAGATT 765

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
 DB 766 CATTTTGACATCAAACTCCCTTGATCATGCTAAATAATGTGATTTTCTTTACTGGA 825

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGluGlyIleGluLys 100  
 DB 826 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAGAGAGATTAAGAG 885

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
 DB 886 GGGCAATCTCTGCTTCGACGACGACATTTGCTGCCACAGTACAGCACAAATGGCATGT 945

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
 DB 946 CATCTTGAAAAAGAACACATCGGCTATTCTGTTTGTAGACAGAGACTTTTACT 1005

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAlaSerAspPheTyrlleArgArg 160  
 DB 1006 CAAATATATGAGATCTGTGTCATCTGTGTGTGTCGAAAGTACTTCTATATCCGACA 1065

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
 DB 1066 GCTCTGGAATTTTAAACAAACGCAACACAGTGCATTTGTGTGTCTCTCCACACTA 1125

QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
 DB 1126 TGCACTGATATGCAATTATGATTCATGAAATGTAAGACTAGTGTGCTG 1185

QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrluProLysCysProLeuGlyValAsp 220  
 DB 1186 GGCATTTTACATGACATGAAAGCATCCGCTATGAAACCAAAATGTCTCTTGAGTAGAC 1245

QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239  
 DB 1246 ATATCAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 1302

RESULT 2  
 AR428809 1526 bp DNA linear PAT 18-DEC-2003  
 LOCUS AR428809  
 DEFINITION Sequence 23 from patent US 6642041.  
 ACCESSION AR428809  
 VERSION AR428809.1 GI:40188595  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 1526)  
 AUTHORS Chen, J., Feder, J.N., Nelson, T.C., Krystek, S.R. and Duclos, F.  
 TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
 JOURNAL Patent: US 6642041-A 23 04-NOV-2003;  
 FEATURES location/Qualifiers  
 1..1526  
 /organism="unknown"  
 /mol\_type="genomic DNA"

Alignment Scores:

Pred. No.: 2,456-113 Length: 1526  
 Score: 1240.00 Matches: 239  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR428809 (1-1526)

QY 1 LeuLeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
 DB 85 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCGCTTCTTGAAAGCTTTGGACATA 144

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgAspLeuSerLeuIleLysHisProGlu 40  
 DB 145 GCACGAGGTGACATGCTTGACAGAGGTGGCAAGAACCTTTCTTAATAAATCAATCCAGG 204

QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
 DB 205 TGCTCCACCATGAGTGTGGGAAAGCCATAGAACCTTTGGCCAAACAGGAATAGATT 264

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
 DB 265 CATTTTGACATCAAACTCCCTTGATCATGCTAAATAATGTGATTTTCTTTACTGGA 324

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGluGlyIleGluLys 100  
 DB 325 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAGAGAGATTAAGAG 384

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
 DB 385 GGGCAATCTCTGCTTCGACGACGACATTTGCTGCCACAGTACAGCACAAATGGCATGT 444

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
 DB 445 CATCTTGAAAAAGAACACATCGGCTATTCTGTTTGTAGACAGAGACTTTTACT 504

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaAlaSerAspPheTyrlleArgArg 160  
 DB 505 CAAATATATGAGATCTGTGTCATCTGTGTGTGTCGAAAGTACTTCTATATCCGACA 564

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
 DB 565 GCTCTGGAATTTTAAACAAACGCAACAGTGCATTTGTGTGTCTCTCCACACTA 624

QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
 DB 625 TGCACTGATATGCAATTATGATTCATGAAATGTAAGACTAGTGTGCTG 684

QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrluProLysCysProLeuGlyValAsp 220  
 DB 685 GGCATTTTACATGACATGAAAGCATCCGCTATGAAACCAAAATGTCTCTTGAGTAGAC 744

QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239  
 DB 745 ATATCAAAAGAGTTGGAGAGCTTCCATTAAGTACCAATTAATAATGAGATA 801

RESULT 3  
 BC011904 1908 bp mRNA linear PRI 23-DEC-2003  
 LOCUS BC011904  
 DEFINITION Homo sapiens O-6-ialoglycoprotein endopeptidase-like 1, mRNA (cDNA  
 clone MGC:20293 IMAGE:4121450), complete cds.  
 ACCESSION BC011904  
 VERSION BC011904.2 GI:40225818  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 REFERENCE 1 (bases 1 to 1908)  
 AUTHORS Strausberg, R.L., Feingold, R.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Scheuen, C.M., Schler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Martusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stalcen, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schaefer, T.E., Brownstein, M.J., Uedlin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McErihan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Weller, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hallyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalek, U., Small, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

2 (bases 1 to 1908)

Strausberg, R. Direct Submission Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

On Dec 19, 2003 this sequence version replaced gi:15080281.

Contact: MGC help desk Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: <http://www.nisc.nih.gov/>

Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)

Akhter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, O.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Taurgeon, C., Vogt, J.B., Walker, M.A., Weethey, K.D., Wiggins, J., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

Series: IRAL Plate: 28 Row: 1 Column: 22.

Location/Qualifiers

1. 1908

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:20293 IMAGE:4121450"

/tissue\_type="Muscle, rhabdomyosarcoma"

/clone\_id="N1H MGC\_17"

/lab\_host="DH10B-R"

/note="vector: pOTB7"

1. 1908

/gene="OSGEP1"

/db\_xref="LocusID:64172"

104. 1348

/gene="OSGEP1"

/codon\_start=1

/product="OSGEP1 protein"

/protein\_id="AAH11904.1"

/db\_xref="GI:15080282"

/db\_xref="LocusID:64172"

/translation="MLITKTAGVFKSKKRVKFLSPNPHPTLFIHKIVLGIETSCDDTAADVNDGNVIGEAHSQTEVRLKTCGIVPAACQAPNRIQIVQALASGVSPSDLSAIAITTKPGIALSLVGVLSFLDVLQGLKPKPFIHMEAHALTRILTKK

VERPPLVLLISGGCHLALVGVSDFLILKSLDIAFGMDLDRVARRLSLIKPEGST  
MSGKAIIEHLAKQGNFHDIDIKPLHANKDCFSFPGIQVTDKILMKKEKEIGK  
GLSSAADIATATQVNMACHIVERTALIFCQRDLIPNNMLVLSGVANPYR  
RLRLITLNTAQCTLLCPRLCTCDNIMIAMNIOEHLRAGLILHIEGRVPEKPL  
GVDISKEVEBASIKVPOLKMET"

215. 1111

/gene="OSGEP1"

/note="Peptidase M22; Region: Glycoprotease family"

/db\_xref="CD:pfam00814"

ORIGIN

Alignment Scores:

Pred. No.: 3,19e-113 Length: 1908

Score: 1240.00 Matches: 239

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC011904 (1-1908)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAaspPheLeuLeuGlnGlySerLeuAasp 20

DB 629 CTGTTGGCATTAGTTCAGAGAGTTTCAGATTCTCTTGGAAAGTCTTGGACATA 688

QY 21 AlaProGlyAaspMetLeuAaspLysValAlaArgArgLeuSerLeuLysHisProGlu 40

DB 689 GACCCAGGAGACATGCTTGACAGAGTGCCAGAGAACCTTCTTAATAAACAATCCAGAG 748

QY 41 CysSerThrMetSerGlyGlyValAlaGlnHisLeuAlaLysGlnGlyAasnArgPhe 60

DB 749 TCGTCCACATGATGCTGGGAAAGCCATAGAACATTGGCCAAACAGGAATAAGTTT 808

QY 61 HisPheAaspIleLysProProLeuHisHisAlaLysAasnCysAaspPheSerThrGly 80

DB 809 CATTTTGACATCAAACTCCCTTCATCATGCTAAAAAATTTGATTTTCTTTTACATGGA 868

QY 81 LeuGlnHisValThrAaspLysIleIleMetLysLysGlnGlnGlnGlyIleGluLys 100

DB 869 CTTCAACACGATGATGATTAATAAATAAGAAAGAAAGAAAGAAAGATATTGAGAAC 928

QY 101 GlnGlnIleLeuSerSerAlaAlaAaspIleAlaAlaThrValGlnHisThrMetAlaCys 120

DB 929 GGGCAAAATCTGCTTTCAGCAGACGACATTGCTCCACAGTACACACAAATGCGATGT 988

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAaspLeuPro 140

DB 989 CATCTTGAAAAGAACACATCGGGCTATCTGTTTGTAAAGCAGAGACTGTAACT 1048

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPhetyrIleArgArg 160

DB 1049 CAATAAATAAGCAGTATGCTGTGCATCGGTGCTGCAAGTAATCTTAATATCCACAGA 1108

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180

DB 1109 GCTCTGGAATTTTAACAAACGACACAGTGCATTTGTTGTCTCTCTCCACAGCTA 1168

QY 181 CysThrAaspAsnGlyIleMetIleAlaArgPasnGlyIleGlnArgLeuArgAlaGlyLeu 200

DB 1169 TGCACGATGAATGCGATTAATGATGCAAGAAAGTATGAAGAAAGTATGAGTGCCTTG 1228

QY 201 GlnIleLeuHisAaspIleGlnGlyIleArgArgGlnProLysCysProLeuGlyValAasp 220

DB 1229 GGCATTTTACATGACATAGAGAGCCATCCCTATGAACCAAAAGTCTCTTGGAGTAGAC 1288

QY 221 HisSerLysGlnValGlnGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239

DB 1289 AATATCAAAAGAAATGGAGAGCTTCATTAAGTAAGTCAAAATTAATAAATGAGAGTA 1345

RESULT 4

AR428803

LOCUS AR428803 2197 bp DNA linear PAT 18-DEC-2003

DEFINITION Sequence 1 from patent US 6642041.

ACCESSION AR428803  
VERSION AR428803.1 GI:40188589  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2197)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 1 04-NOV-2003;  
FEATURES  
Location/Qualifiers  
1..2197  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:  
Pred. No.: 3.77e-113 Length: 2197  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR428803 (1-2197)

QY 1 LeuenuAlaleuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 756 CTGTTGGCATTAGTTCAGAGAGATTTCAGATTTTCCTTGGAAAGCTTTGGACATA 815

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 816 GCACCAAGTGACATGCTTGACAGAGTGCGCAAGAACCTTTCTTAATAAACATCCAGAG 875

QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuValAlaGlnGlyAsnArgPhe 60  
DB 876 TGCTCCACCATGAGTGTGGGAAAGCATAGAACATTTGGCCAAACAGGAATAGATT 935

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
DB 936 CATTTTGACATCAAACTCCCTTGATCATGTCTAAAAATTTGATTTTCTTTACTGGA 995

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100  
DB 996 CTTCAACACGTTACTGATTAATAATTAATGAAGAAAGAAAGAGAGATTTGAAGAG 1055

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
DB 1056 GGGCAAAATCCTGTCTTCAGCAGCAGACATGCTGCGACAGTACAGACACATGGCATGT 1115

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 1116 CATCTTGGAAGAAAGACATGGGCTATTCTGTTTGTAAACAGAGAGACTTGTTACT 1175

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaLysAsnPheThrIleArgArg 160  
DB 1176 CAAATATATGCGATGCTGTCATCTGTGTGTGTCGCAAGTACTTCTATATCCGCA 1235

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
DB 1236 GCTCGGAAATTTTAAACAAAGCAGACACAGTGCATTTGTGTGTCTCTCCAGACAT 1295

QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 1296 TGCACTGTAATGCGATTATGATTCATGAGATGATTTGAAGAGCTAGCTCGCTTG 1355

QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgGluProLysCysProLeuGlyValAsp 220  
DB 1356 GGCATTTTAACTAGACATGAGAGGATCCGCTATAGAACCAAAATGTCCTCTTGAAGAG 1415

QY 221 IleSerLysGluValGlyValAspSerIleLysValProGlnLeuLysMetGluIle 239  
DB 1416 ATATCAAAAGAGAGTGGAGAGCTTCATTAAGATACCAATTTAAATAATGAGAT 1472

RESULT 5  
LOCUS AR428808 1387 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 21 from patent US 6642041.  
ACCESSION AR428808  
VERSION AR428808.1 GI:40188594  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1387)  
AUTHORS Chen,J., Feder,J.N., Nelson,T.C., Krystek,S.R. and Duclos,F.  
TITLE Polynucleotides encoding a novel metalloprotease, MP-1  
JOURNAL Patent: US 6642041-A 21 04-NOV-2003;  
FEATURES  
Location/Qualifiers  
1..1387  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Alignment Scores:  
Pred. No.: 9.37e-111 Length: 1387  
Score: 1213.50 Matches: 238  
Percent Similarity: 90.15% Conservative: 0  
Best Local Similarity: 97.86% Mismatches: 1  
Query Match: 97.86% Indels: 25  
DB: Gaps: 1

US-10-649-273-2\_COPY\_176\_414 (1-239) x AR428808 (1-1387)

QY 1 LeuenuAlaleuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 549 CTGTTGGCATTAGTTCAGAGAGATTTCAGATTTTCCTTGGAAAGCTTTGGACATA 608

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 609 GCACCAAGTGACATGCTTGACAGAGTGCGCAAGAACCTTTCTTAATAAACATCCAGAG 668

QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuValAlaGlnGlyAsnArgPhe 60  
DB 669 TGCTCCACCATGAGTGTGGGAAAGCATAGAACATTTGGCCAAACAGGAATAGATT 728

QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
DB 729 CATTTTGACATCAAACTCCCTTGATCATGTCTAAAAATTTGATTTTCTTTACTGGA 788

QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIle----- 98  
DB 789 CTTCAACACGTTACTGATTAATAATTAATGAAGAAAGAAAGAGATATATTCTTA 848

QY 98 ----- 98

DB 849 ATTATGTAAGTTGAACAGATTAATATTCTGATGTCCTTAATAAATAGCTGCTATTTC 908

QY 99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115  
DB 909 TCGAGGTATGAGAGGGGCAAAATCCTGTCTTCAGCAGACACATTCGCAACAGACAG 968

QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135  
DB 969 CACACATAGCATGTATCTTGAAAGAAACACATCGGCTATTTCTGTTTGTAAAGAG 1028

QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaLysAsn 155  
DB 1029 AGAGACTTGTACTCAAAATATATGCAAGTACGTGTGTGTGTCGCAAGTAAAC 1088

QY 156 PheThrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCys 175  
DB 1089 TTCTATATCCGAGAGACTCTGGAATTTTAAACAAAGCAGACACAGTGCATTTGTGTGT 1148

QY 176 ProProProArgLeuCysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArg 195

Db 1149 CCTCCTCCAGACTATGCACTGATTAATGCAATTGATGCAATGATTTGGAAGA 1208  
QY 196 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyIleArgTyrGlyProIleCys 215  
Db 1209 CTAGCTGGTGGCTTGGGCAATTTTACATGACATGAGGCAATCCGCTATACCAAAATGT 1268  
QY 216 ProLeuGlyValAspIleSerIleGlyValGlyIleValAspIleLeuValProGlyLeu 235  
Db 1269 CCTCTGGAGTAGACATATCAAAAGAGTTGGAAGAGCTTCCATAAAGTACACAAATTA 1328  
QY 236 LysMetGluIle 239  
Db 1329 AAAATGAGATA 1340  
RESULT 6  
HSA295148 1387 bp mRNA linear PRI 30-OCT-2000  
LOCUS HSA295148 Homo sapiens mRNA for putative sialoglycoprotease type 2.  
ACCESSION AJ295148.1 GI:11071726  
VERSION AJ295148.1 GI:11071726  
KEYWORDS metalloproteinase; sialoglycoprotease.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Chen J.M., Fortunato M. and Barrett A.J.  
TITLE Cloning and sequencing of a second human putative  
JOURNAL sialoglycoprotease homologue  
AUTHORS 2 (bases 1 to 1387)  
TITLE Unpublished  
JOURNAL Direct Submission  
AUTHORS Chen J.M.  
TITLE Submitted (27-OCT-2000) Chen J.M., NEC Molecular Enzymology  
JOURNAL Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,  
UNITED KINGDOM  
FEATURES  
source 1. 1387  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="uterus"  
24. 1343  
/function="metalloproteinase of family M22"  
/codon\_start=1  
/product="putative sialoglycoprotease type 2"  
/protein\_id="CAC14666.1"  
/db\_xref="GI:11071727"  
/db\_xref="GOA:O9H4B0"  
/db\_xref="UniProt/TREMBL:O9H4B0"  
/translation="MLILTKTAGVFPFKSKRVYEFLLSPNHPBTLFLAKIVLGIEI  
SCDPTAAVVDENNVLEBAIHSQTEVHLKTKGIYPPAQQLEHRENIORIVQELASAS  
GVSPDLSAIAITTKPGIALSLIGVLSFDLVGQLKRPPIPIHMEALATILRNK  
VEPFVLILISGHCIALVQGVSDFLIGSLDLPQDMLDKVARRLSLIKHPECS  
MSGKALTEHLAKGNRPFPDIKPLHAKNCDSPETGQHYTDKIIMKKKEGFI  
SKPEQINIPGLCTKIAAHFCRYKEGQILISSAADIATYQHTMACILVKTIRAILFCK  
QRDLIPQNNANVLVASGVASNFYIRRALEILLTNAQCTLLCPPLCTDNGIMIAWNG  
IERLRGIGILHIDIEGIRYBKCPGVDISKEVGEASIKVQLKMEI"  
ORIGIN  
Alignment Scores:  
Pred. No.: 9.37e-111 Length: 1387  
Score: 1213.50 Matches: 238  
Percent Similarity: 90.15% Conservative: 0  
Best Local Similarity: 90.15% Mismatches: 1  
Query Match: 97.86% Indels: 25  
DB: 9 Gaps: 1  
US-10-649-273-2\_COPY\_176\_414 (1-239) x HSA295148 (1-1387)  
QY 1 LeuLeuValLeuValGlyIleGlyValSerAspPheLeuLeuGlyIleSerIleAspIle 20  
Db 549 CTGTGGCATTTAGTACAGAGATTTCAGATTTTCGCTTCTTGAGAAAGCTTTGAGACATA 608

QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerIleLeuHisProGly 40  
Db 609 GCACCAAGGACATGCTTTGACAAAGGCGCAAGAAAGCTTTCTTAATAAATCCAGAG 668  
QY 41 CysSerThrMetSerGlyGlyValAlaIleGlyHisLeuValAlaGlyIleAspArgPhe 60  
Db 669 TGCTCCACCATGAGTGTGGGAAAGCCATAGACATTTGGCCAAACAAAGGAAATGATTT 728  
QY 61 HisPheAspIleLeuSerProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
Db 729 CATTTTGACATCAAACTCCCTGCATGCTCAAAATTTGTATTTTCTTTTACTGGA 788  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlyIleGlyIle----- 98  
Db 789 CTTCAACAGCTTACTATTAATAATTAATGAAGAAAGGAAAGGAGATTAATTTCTA 848  
QY 98 ----- 98  
Db 849 ATTAGTAAGTTGAACAGATAATATATTCCTGATGTGCTTAATAAATAGCTGCTATTTC 908  
QY 99 -----GlyLysGlyIleIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115  
Db 909 TGCAGGTATGAGAGGGGCAAAATCCTGTCTTCAGCAGACATGCTGCACAGTACAG 968  
QY 116 HisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135  
Db 969 CACACATGCGCATGTCATCTTGTAAGAAAGAACATCGGGCTATTTCTGTTTGAAGCAG 1028  
QY 136 ArgAspLeuLeuProGlnAsnAlaValLeuValAlaSerGlyIleValAlaSerAsn 155  
Db 1029 AGAGACTTTTACCTCAAAATTAATGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1088  
QY 156 PheTyrIleArgArgAlaLeuGlyIleLeuThrAsnAlaThrGlyIleLeuLeuCys 175  
Db 1089 TTTCTATATCCGACAGAGCTCTGAAATTTTAACAAACGACAGACAGCACTTTGTTGTGT 1148  
QY 176 ProProProArgLeuLeuSerThrAspAsnGlyIleMetIleAlaIleThrAsnGlyIleGlyArg 195  
Db 1149 CCTCCTCCAGACTATGCACTGATTAATGCAATTAATGCAATTAATGCAATTAATGCAATTA 1208  
QY 196 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyIleArgTyrGlyProIleCys 215  
Db 1209 CTAGCTGGTGGCTTGGGCAATTTTACATGACATGAGGCAATCCGCTATACCAAAATGT 1268  
QY 216 ProLeuGlyValAspIleSerIleGlyValGlyIleValAspIleLeuValProGlyLeu 235  
Db 1269 CCTCTGGAGTAGACATATCAAAAGAGTTGGAAGAGCTTCCATAAAGTACACAAATTA 1328  
QY 236 LysMetGluIle 239  
Db 1329 AAAATGAGATA 1340  
RESULT 7  
AX664697 1245 bp DNA linear PAT 22-MAR-2003  
LOCUS AX664697 Sequence 6 from Patent WO02074960.  
ACCESSION AX664697  
VERSION AX664697.1 GI:29164457  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Leiby, K.R., Kapeller-Libermann, R. and Gluckmann, M.  
TITLE 38650, 28472, 5495, 65507, 81368 and 14354 methods and compositions  
JOURNAL of human proteins and uses thereof  
AUTHORS Patent: WO 02074960-A 5 25-SEP 2002;  
TITLE Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source 1. 1245  
/organism="Homo sapiens"

ORIGIN /mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## Alignment Scores:

Pred. No.: 9.1e-110 Length: 1245  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AX664697 (1-1245)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspLe 20
DB 526 CTGTTGGATTAGTTCAGAGGAGTTTCAGATTTTCCTTCTTGAAAGCTTTGGACATA 585
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu 40
DB 586 GCACCAAGTGCACATCTTGACAGAGTGCACAGACACTTTCTTAATTAACATCCAGAG 645
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 646 TGCTCCACATGAGTGGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 705
QY 61 HisPheAspLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 706 CATTTTGACATCAACCTCCCTTGATCATGCTAAATTTGATTTTCTTTACTGGA 765
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
DB 766 CTTCAACACGTTACTGATTAATAATATGAAACACAGAAACAGAGAGCTTTGAGAG 825
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleValIleGlnHisThrMetAlaCys 120
DB 826 GGGCAAAATCCTGCTTTCAGACAGACAGATTTGCTGCACAGTACAGACACAAATGGCAT 885
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 886 CATCTTGGAAGAAACACATCGGGCTATTCTGTTTGTGACAGAGACATTTGTTACT 945
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheThrIleArgArg 160
DB 946 CAATAATATGAGATCTGATGTCATCTGATGTCGAGTAACTTCTTAATCCGACAA 1005
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1006 GCTCTGGAAATTTTAAACAAACGCAACACAGTGCACCTTGTCTGCTCTCCAGACTA 1065
QY 181 CysThrAspAsnGlyLysMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1066 TGCACGTATATGCGATATGATTCATGAGATGATGTAATGAAAGACTACGTCTGCT 1125
QY 201 GlyIleLeuHisAspLysIleGluGlyLysArgGlyLysProLysCysProLeuGlyValAsp 220
DB 1126 GGCATTTTAACTGACATGAGAGGCAATCCGCTATGAAACCAAAATGCTCTTGAAGAGAC 1185
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1186 ATATCAAAAGAGAGTTGGAGAGCTTCCATTAAGATACCAATTAATAATGGAGATA 1242
```

RESULT 8  
AX664695 1820 bp DNA linear PAT 22-MAR-2003  
LOCUS Sequence 4 from Patent WO02074960.  
DEFINITION AX664695  
ACCESSION AX664695.1 GI:29164455  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Leiby, K.R., Kapeller-Libermann, R. and Glucksmann, M.  
TITLE 38650, 28472, 5495, 65507, 81588 and 14354 methods and compositions  
JOURNAL of human proteins and uses thereof  
Patent: WO 02074960-A 4 26-SEP-2002;  
Millennium Pharmaceuticals, Inc. (US)  
Location/Qualifiers

## FEATURES

source

## CDS

1..1820  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
146..1390  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD80044.1"  
/db\_xref="GI:29164456"  
/translation="MLILTKTAGVFPKSKRYVEFLRSFNHPGTLPLHKIVLGIET  
SCDDTAAAVVDEGTGVLEGRHISQTEVHLKTGIVPPAAQOHLRENIORIVQBALAS  
GVSFDSLAIATTKIKGLALSLGVLSFSLQVGLKKEFIPHHMEALTLRLTNK  
VRRPLVILISGHCLIALVOGVSPDLKSLIDLRGMDLVARRLSLIKPEGST  
MGGKRIEHLAKQGRFPRDIKPLHNAKCDSPFGIOHVDKXNNEKQSEIEKG  
QILSGADIAIVQTMACHLVKTRHAILFKQKRDLPQNNAVLVASGVANPPIR  
RALRLTNATVQCTLLCPPERLCTDNGIMTAMNGIERLRAGLGLHDIEGIRYEPKCP  
GVDISKVEGASIKVPLQKMEI"

## ORIGIN

## Alignment Scores:

Pred. No.: 1.43e-109 Length: 1820  
Score: 1203.00 Matches: 232  
Percent Similarity: 98.33% Conservative: 3  
Best Local Similarity: 97.07% Mismatches: 4  
Query Match: 97.02% Indels: 0  
DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AX664695 (1-1820)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspLe 20
DB 671 CTGTTGGATTAGTTCAGAGGAGTTTCAGATTTTCCTTCTTGAAAGCTTTGGACATA 730
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuLysHisProGlu 40
DB 731 GCACCAAGTGCACATCTTGACAGAGTGCACAGACACTTTCTTAATTAACATCCAGAG 790
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 791 TCCTCCACATGAGTGGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 850
QY 61 HisPheAspLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 851 CATTTTGACATCAACCTCCCTTGATCATGCTAAATTTGATTTTCTTTTACTGGA 910
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100
DB 911 CTTCAACAGTTAATCTATTAATAATATGAAACAGAAACAGAGAGGTATTTGAGAG 970
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleValIleGlnHisThrMetAlaCys 120
DB 971 GGCCAAAATCCTGCTTTCAGACAGACATTTGCTGCACAGTACAGACACAAATGGCAT 1030
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 1031 CATCTTGTAAGAAACACATCGGCTATTCGTTTGTGTAAGAGAGAGCTGTATCT 1090
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
DB 1091 CAATAATATGAGATCTGATGTCATCTGATGTCGAGTAACTTCTTAATCCGACAA 1150
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
DB 1151 GCTCTGGAATTTTAAACAAACGACAGTGCATCTTGTGTGTCCTCTCCAGACTA 1210
QY 181 CysThrAspAsnGlyLysMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
```



Db	1211	TGACGTAGTATTCGATTATCATTCATGCATGAATGGTATTTGAAAGCTACGTCTGGCGTGT	1270
Oy	201	G A L E H I S A P I E G U G Y T L A R G Y T G U P O V C G P R O L E U G Y A L A S P	220
Db	1271	GGCAATTTTACATGACATGAAAGGACATCGCATATGAAACAATAATGCTCCCTTGAGAGTAGAC	1330
Oy	221	I L E S E R Y E G U A L G Y G U A L S E R I E Y S V A P P R O G I N L E U Y M E G U I L E	239
Db	1331	ATATCAAAGAAGATTGAGAAAGCTTCCATTAATAAGTACCACATTTAAAATGAGATA	1387
RESULT 9			
BC058172		1844 bp	mRNA linear ROD 08-OCT-2003
LOCUS	Mus musculus cDNA clone MGC:67870 IMAGE:5012054,		complete cds.
DEFINITION			
ACCESSION	BC058172		
VERSION	BC058172.1	GI:3484963	
KEYWORDS	MGC.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE			
AUTHORS	Strausberg,R.L., Fingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschuler,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhac,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heide,F., Diatchenko,L., Marsina,K., Farmer,A.A., Rubin,G.M., Hong,L., Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Udell,T.B., Toshiiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,A., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulik,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettleman,M., Madan,A., Rodrigues,X., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.B., Scherch,A., Schein,J.B., Jones,S.J. and Marra,M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26), 16899-16903	(2002)
MEDLINE	22388257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 1844)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: gcgaps-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMI) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nih.gov		
	Blahet N., Ayele K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Grante, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlin, B., Kwong, P., Latic, P., Legaspi, J.R., Maden, Q.L., Mastello, C., Maeser, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantichop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggin, L., Young, A., Zhang, L.-H. and Green, E.D.		

FEATURES  
source

1. 1844  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CZECH 11"  
/db\_xref="taxon:10090"  
/clone="MGC:67870 IMAGE:5012054"  
/tissue type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MTVV."  
/clone\_1db="NCI CGAP\_Lu29"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
222. 1466  
/codon\_start=1  
/product="unknown (protein for MGC:67870)"  
/protein\_id="AAH58172.1"  
/db\_xref="GI:34849664"  
/translation="MLMLRPTAGAIPKPKSKYVGLRFSVHPRTLSCHKLVGIE  
RCDPTGAAYVDENGVGALSHOTOVHLKTGGVTPPAOOLHENTORIVERTLSA  
RIPSDLSAATATTKKPLGALSLGVGLSDLVNRFKPTPIHMEHALTISLTN  
VEFFVLVLSGGHCLALVGVSDPFLGLSLDLPEDMLDKARRLSLKHPGST  
MSGGAKIEHLAKQGNRFHPTINPPWNAKNCDFPTGQHTDGLTHKEELEGK  
QISSAADIAAAVGNHATACHLAKTRTHALIFCCKNLSPNAAVLVGVASNLVYI  
KALIEIVANQETCLCPPTLCCTDNGIMIANWGIERLRAGLHVLDVEDIRPEKCP  
GIDISREVAAEAIKVPRPKMAL"

misc\_feature  
333. 1397  
/note="ORF17, Region: Metal-dependent proteases with possible  
chaperone activity [Posttranslational modification, protein turnover,  
chaperones]"  
/db\_xref="CDD:C005033"

ORIGIN

Alignment Scores:  
Pred. No.: 6.02e-98 Length: 1844  
Score: 1086.00 Matches: 205  
Percent Similarity: 92.05% Conservative: 15  
Best Local Similarity: 85.77% Mismatches: 19  
Query Match: 87.58% Indels: 0  
DB: 10 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC058172 (1-1844)

OY 1 leuenuhialeuValGInGlyValSerAspPheLeuLeuGlyVLeuSerLeuAspIle 20  
Db 747 CTGTTGGCATTAGTCCAAAGGTTCGATTCCTGCTCTTGGGAATCTTTTGACACTA 806  
OY 21 AlAProGlyAspPheLeuAspLeuValAlaArgArgLeuSerLeuIleValHisProGlu 40  
Db 807 GCGCCAGGCGACATCTTGACAAAGGTGGCAAGAAAGACTTTTTRATTAACATCCAGA 866  
OY 41 CysSerThetSerSergIyVlyValAlleGluHisLeuAlaVlySgInGlyVasnarGhe 60  
Db 867 TGTTCTACATAGATGAGTGGAAAAGCTATAGAACATTTGGCCAAAGACGAATAGATTCC 926  
OY 61 HisPheAspIleVlyProProleuHisHisAlaVlyAsnCyValAspPheSerPheThyGly 80  
Db 927 CATTTTACTATCATCACTCACTATCCAGAAAGTCTAGAAATTCGATTTTCTTTTAAACGGA 986  
OY 81 leuGlnHisValThrAspVlyIleIleMetVlySgInGlyGluGluGlyIleGluVlys 100  
Db 987 CTTCACACATATACCTGATTAAGCTAAATACACACACAAAGAAAAGAAAGGCAATTGAGAG 1046  
OY 101 GTCGInIleLeuSerSerAlaValAspIleValAlaThrValGlnHisThrMetLeuAsyS 120  
Db 1047 GGGCAATTCCTGATCATCAGCGCAGACATTCCTGCTGGCGTACAGCATCAACAGCGGCG 1106  
OY 121 HisLeuValVlysarGlnHisValGluAlleLeuPheCyVlySgInaVgAspLeuLeuPro 140

```

|||||
1107 CACCTTCACAAAACACATCGTGTCTATCTGTTTCAGACAGAAAATTTGCTACT 1166
141 GlnAaenAaValleuValAlaSerGlyValAlaSerAaPheYrTleArg 160
1167 CCAGCTAACGAGATTAAGTGTATCTGAGGTGTGACGAGTGTACATCGAATA 1226
161 AlaLeuGluTleuThraSnaAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
1227 GCATTGGAATTTGCGCAATGCAAGCAGTGCCTTCTGTCTCCCTCCAGACTG 1286
181 CysThraSnaenGlyIleMetIleAlaTrpangGlyIleGluArgLeuArgAlaGlyLeu 200
1287 TGCACTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1346
201 GlyIleLeuHisAaPileGluGlyIleArgTyrGluProGlyCysProLeuGlyValAaP 220
1347 GGCCTTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1406
221 ILeSerIyGluValGlyGluAlaSerIleLeuValProGlnLeuLeuMetGluIle 239
1407 ATATCCAGAGAGATTCAGAAAGCTCCATTAAGTACCGCATTAATAATGACACTT 1463

RESULT 10
LOCUS BC038910 1017 bp mRNA linear ROD 21-OCT-2003
DEFINITION Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (cDNA
ACCESSION BC038910 GI:24433548
VERSION BC038910.1
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus (house mouse)
REFERENCE 1 (bases 1 to 1017)
AUTHORS Klausner,R.D., Collins,F.S., Wagner,L.H., Berge,J.G.,
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marisina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mulhany,S.D., Bosak,S.A., McEwan,P.J.,
McKernan,K.V., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalium,D.E.,
Schnerch,A., Schein,J.B., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
23888257
12477932
2 (bases 1 to 1017)
Strausberg,R.
Direct Submission
Submitted (25-OCT-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: Jeffrey B. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,

```

```

BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McEay, Steven
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalium, Michael Smith, Lorraine Spence, Jeff Scott,
Michael Thorne, Miranada Teal, Natasia van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAX Plate: 86 Row: f Column: 12.
Location/Qualifiers
1..1017
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5053559"
/clisue_type="Liver, normal, 5 month old male mouse."
/clone_1ib="NCI CGAP_L19"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"

ORIGIN
Alignment Scores:
Pred. No.: 1,48e-97 Length: 1017
Score: 1079.00 Matches: 205
Percent Similarity: 91.63% Conservative: 14
Best Local Similarity: 85.77% Mismatches: 20
Query Match: 87.02% Indels: 0
DB: 10 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x BC038910 (1-1017)
1 LeuLeuAlaLeuValGlnGlyValSerAaPheLeuLeuGlyIleYsSerLeuAaPile 20
110 CTGTTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 169
21 AlaProGlyAaPheMetLeuAaPheValAlaArgArgLeuSerLeuIleYsHisProGlu 40
170 GACACAGGCGAGATGTTTACACAGGCGCAAGAGCTTTCTTAATCAACATCCAGAA 229
41 CysSerThmetSerGlyGlyIleValAlaGlnHisLeuAlaIleGlnGlyAaPhe 60
230 TGTCTTACATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 289
61 HisPheAaPileLeuProProLeuHisHisAlaIleAaPheCysAaPheSerPheThrGly 80
290 CATTCTTACTATCAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 349
81 LeuGlnHisValThraSnaPheIleIleMetClyIleGlyGlnGlyIleGlyIle 100
350 CTTCAACATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 409
101 GlyGlnIleLeuSerSerAlaAlaAaPileAlaIleAlaIleValGlnHisThrMetAlaCys 120
410 GGGCAAAATCTGTATCACTGATGATGATGATGATGATGATGATGATGATGATGATG 469
121 HisLeuValIleArgThrHisArgAlaIleLeuPheCysGlyGlnArgPheLeuPro 140
470 CACCTTGCAGAAAGACATGCGCTTATCTGTTTGGACGAGAAATTTGCTCTCT 529
141 GlnAaenAaValleuValAlaSerGlyValAlaSerAaPheYrTleArg 160
530 CCAGCTAACGAGATTAAGTGTATCTGAGGTGTGACGAGTGTACATCGAATA 589
161 AlaLeuGluTleuThraSnaAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
590 GCATTGGAATTTGCGCAATGCAAGCAGTGCCTTCTGTCTCCCTCCAGACTG 649

```

QY	181	CyTTRAspAsnGlyYIleMetIleAlaTrpAsnGlyIleGluValArgLeuAlaGlyLeu	200
DB	650	TGCATGCAATGCGCATCATGATTGCATGGAAATGGAAATTTGAAGATTACGCTCGCTTG	709
QY	201	GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp	220
DB	710	GCGCGTTTAAACAGATGTGTAGAACATCCGATGTGAACCAAAATGTCCTTGGAATTAGC	769
QY	221	IleSerLysGluValGlyGluAlaSerIleLysValProGluIleLeuMetGluIle	239
DB	770	ATATCAAGAAAGTTGCAGAGCTCCATPAAAAGTACCGCATTTAAAAATGGCACTT	826
RESULT 11			
LOCUS	AX713716	2208 bp	DNA
DEFINITION	Sequence 400 from Patent EP1293569.		linear
ACCESSION	AX713716		
VERSION	AX713716.1	GI:29888642	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,O.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Taneichi,K.I., Seki,N., Yoshikawa,T., Otsuka,M., Negahari,K. and Masuoka,Y.		
TITLE	Full-length cDNAs		
JOURNAL	Patent: EP 1293569-A 400 19-MAR-2003; Helix Research Institute (JP) ; Research Association for Biotechnology (JP)		
FEATURES	Location/Qualifiers		
source	1..2208 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"		
ORIGIN			
Alignment Scores:			
Pred. No.:	3,57e-95	Length:	2208
Score:	1059.00	Matches:	211
Percent Similarity:	88.28%	Conservative:	0
Best Local Similarity:	88.28%	Mismatches:	4
Query Match:	85.40%	Indels:	24
DB:	6	Gaps:	1
US-10-649-273-2_COPY_176_414 (1-239) x AX713716 (1-2208)			
QY	1	LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle	20
DB	869	CTGTGGCATTTAGTTCAAGAGATTTCAATTTTCGCTTCTTGAAAGCTCTTGACATA	928
QY	21	AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu	40
DB	929	GCACCAAGGTGACATCTTTCAGACAGGTGCAGAAAGACTTCTTTAATPAAACATCCAGAG	988
QY	41	CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe	60
DB	989	TGCTTCACCAAGAGGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATATGAGATT	1048
QY	61	HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly	80
DB	1049	CATTTGACATCAACCTCCCTTGATCATCTAAAAATTGTGATTTCTTTTACTGGA	1108
QY	81	LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys	100
DB	1109	CTTCAACACGTTACTGATGATTAATATATATGAAAAAGAAAAAGGAAAGTATTGAGAG	1168
QY	101	GlyGlnIleLeuSerSerValAlaAspIleAlaAlaThrValGlnHisPheMetAlaCys	120
DB	1169	GAGCAATCTCTGCTTCAGCAGACATCTCTGCAACAGTACGACACACATGCAATGCAATGT	1228

[illegible]

Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,  
Kamura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,  
Kikuchi, H., Kanda, K., Megetsuna, M., Murakawa, K., Kanehori, K.,  
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,  
Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 2208)  
Isogai, T., Otsuki, T. and Sugiyama, T.  
Direct Submission  
Submitted (24-Oct-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan, cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.  
FEATURES  
Source  
Location/Qualifiers  
1..2208  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="FEBRA2004592"  
/tissue\_type="brain"  
/clone\_id="FEBRA2"  
/dev\_stage="fetus"  
/note="cloning vector: pME18FL3"  
344..1438  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB70923.1"  
/db\_xref="GI:16550167"  
/translation="MLILTKTAGVFEKSRKRYVFRSPNPHPTGLFKHIVGIET  
SCDPTAAVVDRTGVNAGEATHSOTENHKTGGVPPAAQQLHENTORIYOALSAS  
GVSPSDLSAIRTTIKPGALSLVGLSFLDVGQKCPPIPHMEAHALTRITLNK  
VEPFVLVLLISGHCLALVQGVSDFLGLGSLDIPADMDKVARPLKHPCST  
MSGGKAIHLAKQGNRFHFDIKPLHAKNCDPFTGLOHTYDKIMKEKEGIEKG  
QILSSADIAITVOHMACHLVKRTTHRAILFCCKRDLLPNNNAVLVASGVASNFCIR  
RALEILNATGCTILCPPRCTDNGIMIA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.57e-95 Length: 2208  
Score: 1059.00 Matches: 211  
Percent Similarity: 88.28% Conservative: 0  
Best Local Similarity: 88.28% Mismatches: 4  
Query Match: 85.40% Indels: 24  
DB: 9 Gaps: 1  
US-10-649-273-2\_COPY\_176\_414 (1-239) x AK055441 (1-2208)  
QY 1 LeuleuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 869 CTGTTGGCATTAGTTCAGAGAGTTCAGATTTTCGCTTCCTTGGAAAGCTTTGACATCA 928  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 929 GCACCAAGTGCAGAGCTTGACAGAGTGCAGAGAGCTTCCTTAATAAATCATCCAGAG 988  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 989 TGCTCCACCATGAGTGGTGGGAAGCCATAGACATTTGGCCAAACAGAAATGATTTT 1048  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80  
DB 1049 CATTTTACATCAAACTCCCTTCATCATCTATAAATTTGATTTCTTTTACTGGA 1108  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlyLysGlnGlnGlyIleGlnLys 100

DB 1109 CTTCAACACGTTACTGATTAATAATTAATGAAGAAAAGAGAGAGGATATTAGAGAG 1168  
QY 101 GYGINLILEUSETSERIALAASPILEAIAAETHVALGIMHISPHMETALACYS 120  
DB 1169 GGGCAAACTCTCTTCAGACAGACAGATTCGTCACAGTACAGACCAATGGAGATGT 1228  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 1229 CATCTTGCAAAAAGAACACATCGGCTATTCTGTTTGTAGCAGAGACATCTTTACT 1288  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnAspPheTyrIleArg 160  
DB 1289 CAAATTAATGACGACTAGTGGTTCATCGTGGTGGTGGCAAGTAATCTGTATCCGACA 1348  
QY 161 AlaLeuGlnLileuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlu 180  
DB 1349 GCTCTGGAATTTTAAACAACCAACACAGTGCATCTTGTGTCTCTCCACACTA 1408  
QY 181 CysThrAspAsnGlyIleMetIleAlaTyrAsnGlyIleGluArgLeuArgIleu 200  
DB 1409 TGCATCTGATTAATGGCATTAATGATTGCA----- 1435  
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220  
DB 1436 -----TGAGTCTCTTGGAGTAGAC 1456  
QY 221 HESERTLYSGILVALIGLYLALASERTILEYVALPROGLINLEULYSMETGLUTLE 239  
DB 1457 AATTCAAAAGAAAGTGGAGAACCTTCATTAAGTACCAATTAATAATGAGATA 1513  
RESULT 13  
BC078974 1546 bp mRNA linear ROD 03-AUG-2004  
LOCUS BC078974  
DEFINITION Rattus norvegicus cDNA clone IMAGE:7111906, partial cds.  
ACCESSION BC078974  
VERSION BC078974.1 GI:50926879  
KEYWORDS  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 1546)  
AUTHORS Strusberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shemmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, D., Hsieh, F.,  
Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schetz, T.E., Brownstein, W.J., Usdin, T.B., Toshiyuki, S.,  
Carrinoci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hultky, S.M.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Faney, D., Helton, B., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, D.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Kravinsky, M.I., Skalska, U., Smalins, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
JOURNAL PUMED 12477932  
PUBMED 2 (bases 1 to 1546)  
12477932  
DIRECTOR MGC Project.  
AUTHORS  
TITLE  
JOURNAL  
SUBMITTED (02-AUG-2004) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

## COMMENT

Contact: MGC help desk  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Howard Jacobs  
CDNA Library Preparation: Express Genomics  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 182 Row: f Column: 6  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

## FEATURES

## source

1. 1546  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="IMAGE:7111906"  
/tissue\_type="Testis, rat (Brown Norway)"  
/clone\_id="N1H\_MGC\_237"  
/lab\_host="DH10B"  
/note="Vector: pExpress1"  
144. >1546  
/codon\_start=1  
/product="Unknown (protein for IMAGE:7111906)"  
/protein\_id="AAH78974.1"  
/db\_xref="GI:50926889"  
/translation="MLMSKTAGAI PRPNSNVGFI RRPVQPRALFHHKLVGIET  
SCDPAAVVDETVNGEALSHQTEVHLKTGCVLPVPAQQLHRENIQRIVEALSAS  
GVSPDLSAATTKPGIALSLGVSFVQVLPVQKRPPIHMEALALIRLTHK  
VGPFVLILISGHCIALVQSVDFLLGSLDAPGMDLKVRRSLILKHPECS  
MSGKALEHAKGRRPHFTNPPMNANKDPSFGVQHTDKLITHEKEGEBK  
QIISSAADIAAVQHTACHLAKTTHAILTCQANLSPANAVALVSGVASNLYIR  
RALEIVANATQCTLCPPRLCTGNGIMIANNGERLAGGILADVDDIYEPVAGE  
IVSWRLTALTELDISVSTHTVAHSPINSGSRGANIQTSMCSCTQTVNMTVHTL  
NINERKSKKKKKKKK"

## CDS

ORIGIN  
Alignment Scores:  
Pred. No.: 8.27e-88 Length: 1546  
Score: 983.00 Matches: 188  
Percent Similarity: 92.06% Conservative: 9  
Best Local Similarity: 87.85% Mismatches: 17  
Query Match: 79.27% Indels: 0  
DB: 10 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC078974 (1-1546)

QY 1 leuleuValleuValgInglValSerAspHeuLeuGlyLysSerLeuAspIle 20  
DB 669 CTGTGGCGTAGTTCAGAGTGTTCAGATTTCCTGCTCCGCGAAGTCCCTGACATA 728  
QY 21 AlaProGlyAspMetLeuAspIleValAlaArgArgLeuSerLeuIleHisProGlu 40  
DB 729 GCGCAGGCGACATGCTTGACAAAGTGGCAAGACCTTTTATATCAACATCCAGAA 788  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 789 TGTTCTACATGAGTGTGGGAAAGCTATAGACATTTGGCCAAAGAAAGAAATGATTC 848  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGly 80  
DB 849 CACTTACTACTCATCATCCACCATCATGAGAACTGATTTCTTTTAACTGCGGA 908  
QY 81 leuGlnHisValThrAspIleIleIleMetLysGlnGlyLysGlnGlnGlyTleGlnLys 100  
DB 909 CTTCACATGTCACGATGAGTAAATACACAAAGAAAGAAAGCAATGAGGAAG 968

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120

DB 969 GGGCAAACTCTGTCATCAGCCGACGACATTCCTGCTCGGTACAGCAGCAACAGCGTGC 1028

QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140

DB 1029 CACCTTGCGCAAAAGAACATCGTGTATTCGTTTGGCAGCAAGAAATTTGCTACTC 1088

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160

DB 1089 CCAGCTAACGACATATTAGTGTGCTGAGAGGTGTGCAAGTAACCTGTACATCCGAGA 1148

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180

DB 1149 GCATTGAAATGTACCAATGACACACATGACTTTGTTGTCCTCCGAGACTG 1208

QY 181 CysThrAspAsnGlyIleMetIleAlaTyrPasnGlyIleGluArgLeuArgAlaGlyLeu 200

DB 1209 TGCACAGACATGATGATCATGATTCATGAGAAATGAAAGATTACCTGCTGCTTG 1268

QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGlnProLys 214

DB 1269 GCATTTCATCATGATGTAGAGACATCCGATACGAACCAAG 1310

RESULT 14

BC051211 1109 bp mRNA linear ROD 15-APR-2003

LOCUS BC051211 Mus musculus, clone IMAGE:1327545, mRNA.

DEFINITION BC051211

ACCESSION BC051211.1 GI:29881634

VERSION BC051211.1 GI:29881634

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Mus musculus

REFERENCE Burykova; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

JOURNAL 1 (bases 1 to 1109)

Strausberg, R.

Direct Submission

Submitted (14-APR-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

CDNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: amadan@systemsbiology.org

Anup Madan, Jessica Fahy, Erin Helton, Mark Kerteman, Anuradha

Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAX Plate: 113 Row: b Column: 1

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

FEATURES

source

1. 1109

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:1327545"

/tissue\_type="Thymus gland, mouse"

/clone\_id="Soares\_thymus\_2nbMT"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac"

ORIGIN

## Alignment Scores:

Pred. No.: 3,72e-84 Length: 1109  
 Score: 944.50 Matches: 183  
 Percent Similarity: 82.52% Conservative: 20  
 Best Local Similarity: 74.39% Mismatches: 32  
 Query Match: 76.17% Indels: 11  
 Gaps: 2

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC051211 (1-1109)

```

QY      2 LeuAlaIeuValGInglYValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
      148 TTGAAGCTGACTGGTTCATATGAGGCTTATCTTCTTAAT-----AATTGGCAG 198
      22 ProGlyAspMetLeuAspLys-----ValAlaArgArgLeu 33
      199 CCAAAATATATTTATGAAAAAATGTAATGCTGTTTAAATAGGTGGCAGAAAGACTT 258
      34 SerLeuIleLysHisProGlyCysSerTherMetSerGlyGlyValAlaIleGluHisLeu 53
      259 TCTTATTCACAAATCCAGATGTTCTTACATGATGATGATGATGATGATGATGATGATG 318
      54 AlaLysGInglYAsnArgPheHisPheAspIleLysProProLeuHisAlaLysAsn 73
      319 GCCAAAGACGGAATATGATTCATTTTACTATTCATTCACCTATGACAGATCTAAGAT 378
      74 CysAspPheSerPheTherGlyLeuGlnHisValThrAspLysIleIleMetLysGlu 93
      379 TGCATATTTTCTTTCACCGGACTTCACATATTAATGATTAATACACACAGGAA 438
      94 LysGluGluGlyIleGluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaThr 113
      439 AAAGAAAGAGCATTTGAGAAAGGCGCAATCTGTATATCAGCTGCAGACATTCCTCTCG 498
      114 ValGlnHisThrMetAlaCysHisLeuValLysArgThrHisArgAlaIleLeuPheCys 133
      499 GTACAGCATGCAACAGCGGTGCACCTTGCAAAAGAACACATCGCGTATTTGTTTGG 558
      134 LysGlnArgAspLeuLeuProGlnAsnAlaValIleValAlaSerGlyGlyValAla 153
      559 AAGCAAGAAAAATTTGCTCTCTCCAGCTTAACGAGTATAGTTGATGAGGTGTGCA 618
      154 SerAspPheTrpIleArgArgAlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeu 173
      619 AGTACCTTGATCATCCGAAAGCATGGAATGTGCGAAATGCAACGAGCGACGTTG 678
      174 LeuCysProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIle 193
      679 TTGTGTCACTCCAGACCTGTGCACTGCAATGSCATCATGATTCATGCAATGCAAT 738
      194 GluArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTrpGluPro 213
      739 GAAAGATTATTCCTGCTGCGCTGCTTTTACATGATGATGATGATGATGATGATGAT 798
      214 LysCysProLeuGlyValAspIleSerLysGluValGlyGlnAlaSerIleLysValPro 233
      799 AAATGTCTCTTGAAGTACATATCCAGAAAGTTCGAGGAGTCCCATTAAGATACCG 858
      234 GlnLeuLysMetGluIle 239
      859 CGATTAAAAATGCGACTT 876
  
```

RESULT 15

LOCUS BX930694 860 bp mRNA linear VRT 30-MAR-2004  
 DEFINITION Gallus gallus finished cDNA, clone CHEST967014.  
 ACCESSION BX930694  
 VERSION BX930694.2 GI:46016745  
 KEYWORDS Gallus gallus (chicken)  
 SOURCE Gallus gallus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

## REFERENCE

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E.,  
 Crompton, M.D.R., Davies, R.M., Francis, M.D., Grafton, D.V.,  
 Hubbard, S.J., Humphrey, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,  
 Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,  
 Tickle, C. and Wilson, S.A.  
 Direct Submission

## TITLE

Submitted (29-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: chickens@ems.umbist.ac.uk

## COMMENT

On Apr 1, 2004 this sequence version replaced gi:41631222.  
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA  
 sequencing project.  
 This sequence is from the  
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,  
 from a library constructed by Elizabeth Bosch. cDNA was prepared  
 from RNA extracted from limbs, normalised, and poly A-tailed.  
 EcoRI-NotI cut cDNA was then ligated into the vector. Vector:  
 pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI Host: Escherichia  
 coli DH10B.

## FEATURES

source  
 1..860  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST967014"  
 /clone\_11b="CSQCCHN59"  
 /dev\_stage="stage 36"

## ORIGIN

## Alignment Scores:

Pred. No.: 7.86e-71 Length: 860  
 Score: 809.00 Matches: 151  
 Percent Similarity: 79.66% Conservative: 37  
 Best Local Similarity: 63.98% Mismatches: 48  
 Query Match: 65.24% Indels: 0  
 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BX930694 (1-860)

```

QY      1 LeuLeuAlaLeuValGInglYValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
      94 ATCTTGCGACGTGACAGAGAGATTTCAGATTCCTTCTCTGACAGATCCATAGATATA 153
      21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
      154 GCACCAAGTGAATGCTGTGATTAAGTACAGAAAGGCTCTTTAGTGAAGCACCCGAG 213
      41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnIleYAsnArgPhe 60
      214 TGCCACAGCATGCGCGGGGGAAGCAATAGAGCACCTGCTCAAAACCGAGACTGGCAA 273
      61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
      274 CAGTACACTTTCAGACTTCCTCCATGCAACAGTATCGTATCTGATTTCTTCTCCGGA 333
      81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlnIleGluLys 100
      334 CTTGAGAGCCTTGTCACAAAGCCATTTCTGAAAGAAAGAAAGAAAGATATTCAAGA 393
      101 GlnGlnIleLeuSerSerAlaAlaAspIleAlaIleValGlnHisThrMetAlaCys 120
      394 GGGGAATCTCTGCTGCTTAAGACATCGCTGCTGTCACAGACCTAGTGGCTGCT 453
      121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
      454 CATATATTCACCGGACACACCGAGCATCTTCTGATGAAAGAAACGATATATTATA 513
      141 GlnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheTrpIleArgArg 160
      514 CCAAAAATCGCAACTGTTGTATTCAGAGAGAGATTGCAAGTATCATGATATCAGAAAA 573
  
```

```

QY 161 AlaleuGlulIleuThraSnaIatHrgInCystrLeuLeuCySPProProArgLeu 180
Db 574 GGAAGTGCAGACTCTGCAAAATGCAAAAGCGTTTTCCTTCTGCTCTCTCCAAAGCTG 633
QY 181 CysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db 634 TGCACCGATAATGGTGTATGATGATGCAATGGCATTGAAAGGTTGCGTCAGATGT 693
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
Db 694 GGAATTTATACAGTACTGATGGCACCCTACGAAACCAAAAGCTCCCTTGGAATTGAT 753
QY 221 IleserLysGluValGlyGluAlaSerIleLysValProGlnLeuLys 236
Db 754 ATTTCAAAAGAGTGAAGAGGATTCCATCAAGTCCCAAGACTAAG 801

```

Search completed: February 16, 2005, 18:03:04  
 Job time : 3816.22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:09:32 ; Search time 455.139 Seconds  
(without alignments)  
3108.540 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVQGVSDFLILGKSLDI.....DISKEVGEASIKVPLKMEI 239

Scoring table:  
BIOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues  
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+P2n.model DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10649273/runat\_14022005\_114702\_16389/app\_query.fasta\_1.1429  
-DB=N\_Geneseq\_16Dec04 -QPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=40  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=BIOSUM62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273\_QCGN\_1.1\_1057@runat\_14022005\_114702\_16389 -NCPu=6 -ICPU=3  
-NO MMAP -LARGEQUERRY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_16Dec04:  
1: geneseq1980s:  
2: geneseq1990s:  
3: geneseq2000s:  
4: geneseq2001as:  
5: geneseq2001bs:  
6: geneseq2002as:  
7: geneseq2002bs:  
8: geneseq2003as:  
9: geneseq2003bs:  
10: geneseq2003cs:  
11: geneseq2003ds:  
12: geneseq2004as:  
13: geneseq2004bs:  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	1416	8	ABX70950 Novel hum
2	1240	100.0	1526	6	ABs76639 DNA encod
3	1240	100.0	2058	6	ABa93268 Human O-s
4	1240	100.0	2197	6	ABs76635 DNA encod
5	1240	100.0	2572	8	ABr23207 Human pro

6	1203	97.0	1820	6	AAU46856	AAU46856 Human gly
7	1203	97.0	1820	8	ACA60887	ACA60887 Human cDN
8	1203	97.0	1821	10	ABs57020	ABs57020 cDNA enco
9	1059	85.4	2208	10	ADAs2832	ADAs2832 Human cod
10	1059	85.4	2890	12	ADQ24627	ADQ24627 Human sot
11	995.5	80.3	3358	10	ADB31345	ADbq75508 Human dia
12	870	70.2	1572	6	ABQ75508	ABQ75508 Murine s1
13	599	48.3	2734	5	AA584622	AA584622 DNA encod
14	468	37.7	371	12	ADL86725	ADL86725 DNA up-re
15	468	37.7	371	12	ADL86726	ADL86726 DNA up-re
16	332.5	26.8	1601	4	ABU24633	ABU24633 Drosophil
17	332.5	26.8	3656	4	ABU24632	ABU24632 Drosophil
18	332	26.0	1557	3	ACA38454	ACA38454 Arabidops
19	257	20.7	1146	8	ACA26804	ACA26804 Prokaryot
20	248.5	20.0	1000	6	ABP91424	ABP91424 Moraxella
21	248.5	20.0	1000	6	ABK37804	ABK37804 DNA seque
22	248.5	20.0	1044	8	ACA39102	ACA39102 Prokaryot
23	248.5	20.0	94750	4	AA28551	AA28551 Genomic f
24	247	19.9	936	8	ACA20445	ACA20445 Prokaryot
25	247	19.9	1053	12	ADL03120	ADL03120 DNA encod
26	241	19.4	1032	8	ACA43173	ACA43173 Prokaryot
27	239	19.3	4360	6	ABU48239	ABU48239 Eritrichia
28	235.5	19.0	1092	6	ABQ90383	ABQ90383 M. capsul
29	234	18.9	300	3	AAQ00934	AAQ00934 Human col
30	232	18.7	676	4	AAH08019	AAH08019 Human cDN
31	232	18.7	1385	4	AAH15110	AAH15110 Human cDN
32	231	18.6	1026	4	AA554064	AA554064 Pseudomon
33	231	18.6	1026	8	ACA42146	ACA42146 Prokaryot
34	231	18.6	1026	10	ADG73341	ADG73341 P aerugin
35	231	18.6	1059	11	ABD02280	ABD02280 Pseudomon
36	231	18.6	1206	11	ABD02197	ABD02197 Pseudomon
37	229	18.5	1026	10	ADG73343	ADG73343 P aerugin
38	224	18.1	1029	4	AA553309	AA553309 Haemophil
39	224	18.1	1029	8	ACA34150	ACA34150 Prokaryot
40	224	18.1	11000	2	AA742063	AA742063 (6 of
41	217	17.5	9967	13	ADT05493	ADT05493 Haemophil
42	217	17.5	85814	13	ADT05644	ADT05644 Haemophil
43	214	17.3	1044	4	AB108591	AB108591 Drosophil
44	213	17.2	1014	4	AA556045	AA556045 Salmonell
45	213	17.2	1014	8	ACA51431	ACA51431 Prokaryot

ALIGNMENTS

RESULT 1	ABX70950	standard; cDNA, 1416 BP.
ID	ABX70950	
XX	ABX70950;	
AC	ABX70950;	
DT	05-MAR-2003	(first entry)
XX		
DE	Novel human cDNA sequence #175.	
XX		
KW	Human; gene; ss; nervous system disorder; peripheral neuropathy;	AAU46856 Human gly
KW	Huntington's disease; amyotrophic lateral sclerosis; haemophilia;	ACA60887 Human cDN
KW	neurodegenerative disease; Parkinson's disease; Alzheimer's disease;	ABs57020 cDNA enco
KW	autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;	ADAs2832 Human cod
KW	insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;	ADQ24627 Human sot
KW	ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;	ADbq75508 Human dia
KW	fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;	AA584622 DNA encod
KW	coagulation disorder; cancer; tumour; inflammatory disease; septic shock;	ADL86725 DNA up-re
KW	Crohn's disease; anaphylaxis; proliferation; chemoclastic;	ADL86726 DNA up-re
KW	differentiation; stem cell growth factor; haematopoiesis; chemokine;	ABU24633 Drosophil
KW	haemostatic; antiinflammatory; expressed sequence tag; EST.	ABU24632 Drosophil
OS	Homo sapiens.	ACA38454 Arabidops
XX		ACA26804 Prokaryot
PN	WO200281731-A2;	ABP91424 Moraxella
XX		ABK37804 DNA seque
PD	17-OCT-2002.	ACA39102 Prokaryot
XX		AA28551 Genomic f
XX		ACA20445 Prokaryot
PF	29-JAN-2002; 2002WO-US001222.	ADL03120 DNA encod



XX 30-JAN-2001; 2001US-00774528.  
XX (HYSE-) HYSEQ INC.  
XX (GOOD/) GOODRICH R W.  
XX  
XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F,  
XX Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI; 2003-058563/05.  
XX  
XX Novel polypeptide useful for treating neurodegenerative diseases, myeloid  
XX or lymphoid cell disorders, bone disorders, mechanical and traumatic  
XX disorders, coagulation disorders, and inflammatory diseases.  
XX  
XX Claim 1; Page: 612pp; English.  
XX  
XX This invention relates to the cDNA sequences encoding an isolated novel  
XX human polypeptide. The protein encoded by the nucleic acid of the  
XX invention is useful for treating central and peripheral nervous system  
XX diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic  
XX lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,  
XX Alzheimer's disease); autoimmune disease (e.g. systemic lupus  
XX erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)  
XX ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopenia)  
XX ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,  
XX osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head  
XX trauma); lung or liver fibrosis; reperfusion injury in various tissues;  
XX bacterial, viral or fungal infections; allergic conditions such as  
XX allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);  
XX cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's  
XX disease, anaphylaxis). The protein may be used to inhibit the growth,  
XX infection or function of infectious agents such as bacteria, fungi,  
XX viruses, or to effect bodily characteristics, biorythms or circadian  
XX cycles of rhythms. The protein may also have  
XX proliferation/differentiation, stem cell growth factor, haematopoietic,  
XX regulation, immune stimulating or suppressing, chemotactic/chemokinetic,  
XX haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
XX activities. The cDNA sequences of the invention are useful for expressing  
XX recombinant protein for analysis. The present sequence represents a novel  
XX human cDNA sequence of the invention, this sequence is an expressed  
XX sequence tag (EST) and was identified using subtractive hybridisation  
XX  
XX Sequence 1416 BP; 441 A; 280 C; 273 G; 422 T; 0 U; 0 Other;  
XX  
XX Alignment Scores:  
XX Pred. No.: 3-21e-132 Length: 1416  
XX Score: 1240.00 Matches: 239  
XX Percent Similarity: 100.00% Conservative: 0  
XX Best Local Similarity: 100.00% Mismatches: 0  
XX Query Match: 100.00% Indels: 0  
XX Gaps: 0  
XX  
XX US-10-649-273-2\_COPY\_176\_414 (1-239) x ABX70950 (1-1416)  
XX  
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 586 CTGTTGGCATTAAGTTCAGAGATTTCAGATTTCGCTTCCTGGAAGCTTTGACATA 645  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgGlyLeuSerLeuIleLysHisProGlu 40  
DB 646 GCACCAAGGTGACATGCTTGAAGGTGGAGAGACCTTCTTAATATAACATCCAGAG 705  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyLysAsnArgPhe 60  
DB 706 TGCCTCCACCATGAGTGTGGGAAACCATAGAACATTTCGCCAAACAGAGAAATGATTT 765  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
DB 766 CATTTTGACATCAAACTCCCTTCATCATGCTAATAAATTTGATTTTCTTTACTGGA 825  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGlnGluGlyIleGlnLys 100  
|||||

DB 826 CTTCAACACGTTACTGATTAATAATATGAAAAAGAAAAAGAGAGATTTAGCAAG 885  
QY 101 GtGlnIleLeuSerSerIleAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120  
DB 886 GGGCAAAATCTCTCTTACAGACAGACATGTGCTGCACAGTACAGACCAATGSGATGT 945  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 946 CATCTTGAAAAGAAACACATGGGCTATTCTGTTTGTAGCAGAGACATTTGTAACCT 1005  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnAspThrIleArgArg 160  
DB 1006 CAAATTAATGACGTAATGCTGTCATCTGCTGGTGGTGGCAAGTAATCTTAATCCGACA 1065  
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlu 180  
DB 1066 GCTCTGGAATTTTAAACAAACCAACAGTCACTTGTGTGCTCTCCCAACATGA 1125  
QY 181 CysThrAspAsnGlyIleMetIleAlaTPaNGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 1126 TGCACGTATATGCGATTATGATTGCAATGAAATGATGAAAGACTACGTGCTGCTTG 1185  
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrgLupProLysCysProLeuGlyValAsp 220  
DB 1186 GGCATTTTACATGACATAGAGGATCCGCTATGAACCAAAATGTCTCTTGAGTAGAC 1245  
QY 221 HisSerLysGluValGlyGlnAlaSerIleLysValProGlnLeuLysMetGluIle 239  
DB 1246 AATATCAAAAGAAAGTTGGAGAGACCTTCATTAAGTCAATTAATAATGAGATA 1302  
RESULT 2  
ABX76639  
ID ABX76639 standard; DNA; 1526 BP.  
XX  
XX ABX76639;  
AC  
XX  
XX 11-DEC-2002 (first entry)  
DT  
XX  
XX DNA encoding novel human metalloprotease WPI fragment #1.  
DS  
XX  
XX Metalloprotease; MP-1; immune disorder; glutamate transport; cancer;  
KW motor neuron disorder; amyotrophic lateral sclerosis; ALS; diabetes;  
KW reproductive disorder; Klinefelter's syndrome; germinal cell aplasia;  
KW genital wart; metabolic disorder; premature puberty; Kallman syndrome;  
KW Cushing's syndrome; neurodegenerative disease; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; Tourette syndrome; sepsis;  
KW liver disease; renal disease; immune disorder; rheumatoid arthritis;  
KW acquired immunodeficiency syndrome; AIDS; pulmonary disease; pneumonia;  
KW emphysema; cystic fibrosis; vascular disorder; inflammatory disorder;  
KW neurological disorder; gene; de.  
XX  
XX Homo sapiens.  
OS  
XX  
XX MO200272751-A2.  
PN  
XX  
XX 19-SEP-2002.  
PD  
XX  
XX 05-FEB-2002; 2002WO-US003353.  
PF  
XX  
XX 05-FEB-2001; 2001US-0266518P.  
PR  
XX  
XX 10-APR-2001; 2001US-0282814P.  
PR  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
PA  
XX  
XX Chen J, Feder J, Nelson TC, Duclos F, Krystek S;  
PI WPI; 2002-723329/78.  
DR P-PSDB; ABG96487.  
XX  
XX New isolated nucleic acid encoding MP-1 protein, useful for preventing,  
PT treating, or ameliorating diseases associated with aberrant  
PT metalloproteinase activity, e.g. immune, metabolic, inflammatory and  
PT neurological disorders.







QY 161 AAlaLeuGluIleuThrAsnAlaThrGlnCysThrLeuLeuCyseProProArgLeu 180  
 Db 1149 GCTCTGGAATTTTAAACAAGCAACAGTGCCTTTGTGTGCTCTCCAGACTA 1208  
 QY 181 CysThrAspAnglyIleMetIleAlaTPanglyIleGluArgLeuArgAlaGlyLeu 200  
 Db 1209 TGCACGTATATATGCAATATATATGCAATGGAATGATTTGAAGAAGCTACGTCTGGCTTG 1268  
 QY 201 GAlIleLeuHisAspIleGluGlyIleArgTyrGluProLysCyseProLeuGlyValAsp 220  
 Db 1269 GGCATTTTACATGACATAGAAAGCAATCCGCTATGAACCAAAATGCTCTTGGAGTAAAC 1328  
 QY 221 ILeSerIySGluValGlyGluAlaSerIleIySValProGlnLeuIyMetGlyIle 239  
 Db 1329 ATATCAAAAGAGTGGAGAGGCTTCCATTAAGTACCAATTAATAATGAGATA 1385

RESULT 6  
 AAD46856  
 ID AAD46856 standard; cDNA; 1820 BP.  
 XX  
 AC AAD46856;  
 DT 27-JAN-2003 (first entry)  
 XX  
 DE Human glycoprotease 28472 cDNA.  
 XX  
 KW Human; adenosine deaminase; seven transmembrane domain receptor; cancer;  
 KW 7TM; glycoprotease; immune disorder; IGA deficiency; allergy; arhythmia;  
 KW rheumatoid arthritis; diabetes; atherosclerosis; cardiovascular disorder;  
 KW hypertension; ischaemic heart disease; obesity; myocardial infarction;  
 KW endothelial cell disorder; Grave's disease; psoriasis; brain disorder;  
 KW Parkinson's disease; Alzheimer's disease; haematopoietic disorder;  
 KW cerebral oedema; metabolic disorder; liver disorder; platelet disorder;  
 KW chromosome mapping; tissue typing; gene therapy; neuroprotective;  
 KW cytosolic; anorectic; cardiac; haemostatic; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Location/Qualifiers  
 FT CDS  
 FT 146..1390  
 FT /\*tag= a  
 FT /product= "Human 28472 protein"  
 FT /note= "This region is specifically claimed as SEQ ID NO:  
 FT 6 in claim 1 of the specification"

XX  
 PN WO200274960-A2.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 08-NOV-2001; 2001MO-US051427.  
 XX  
 PR 08-NOV-2000; 2000US-0246768P.  
 PR 08-NOV-2000; 2000US-0246772P.  
 PR 15-NOV-2000; 2000US-0249185P.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Ielby KR, Kapeller-Libermann R, Gluckmann M;  
 XX  
 DR WPI; 2002-759898/82.  
 DR P-PDB; AAE29234.  
 XX  
 PT New 38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid molecules,  
 PT useful for diagnosing and treating cancer, immune, cardiovascular,  
 PT hematopoietic, brain, pain, metabolic, liver or platelet disorders, and  
 PT in pharmacogenomics.  
 XX  
 PS Claim 1, Fig 8, 178pp; English.  
 XX  
 CC The present invention relates to novel 38650, 28472, 5495, 65507, 81588  
 CC or 14354 nucleic acid molecules which encode adenosine deaminase, glyco-  
 CC protease or seven transmembrane domain (7TM) receptor family members.  
 CC Sequences of the invention are useful in diagnosing and treating cancer

CC or aberrant cellular proliferation and/or differentiation (e.g. colon or  
 CC lung cancer), immune disorders (e.g. selective IGA deficiency, rheumatoid  
 CC arthritis, diabetes, allergy), heart or cardiovascular disorders (e.g.,  
 CC hypertension, atherosclerosis, arrhythmias, ischaemic heart disease,  
 CC myocardial infarction, thrombus) including endothelial cell disorders  
 CC (e.g. psoriasis, Grave's disease), haematopoietic disorders, brain  
 CC disorders (e.g. cerebral oedema, Parkinson's or Alzheimer's disease),  
 CC pain and metabolic disorders (e.g. obesity), liver disorders or platelet  
 CC disorders. They are also useful in screening assays, predictive medicine  
 CC (e.g. diagnostic assays, prognostic assays, monitoring clinical trials  
 CC and pharmacogenetics) and prophylactic and therapeutic methods. The  
 CC nucleic acids may also be used in chromosome mapping, tissue typing and  
 CC forensic biology and as surrogate markers. Sequences of the invention are  
 CC also used in gene therapy. The present sequence is human glycoprotease  
 CC 28472 cDNA  
 XX  
 SQ Sequence 1820 BP, 543 A, 365 C, 393 G, 518 T, 0 U, 1 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 8-22e-128 Length: 1820  
 Score: 1203.00 Matches: 232  
 Percent Similarity: 98.33% Conservative: 3  
 Best Local Similarity: 97.07% Mismatches: 4  
 Query Match: 97.02% Indels: 0  
 DB: 6 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AAD46856 (1-1820)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyIySerLeuAspIle 20  
 Db 671 CTTGTTGGCATTTAGTTCAAGAGATTTTCAAGATTTTCTGCTTGGAAAGCTTTGGACATA 720  
 QY 21 AlaProGlyAspMetLeuAspIySValAlaAlaGluLeuSerLeuIleIySHisProGlu 40  
 Db 721 GCACCAAGGTGACATCTGTCGACCAAGGTGGCAAGACCTTTCTTAATAAATCCACAGAG 790  
 QY 41 CysSerThrMetSerGlyGlyIySValAlaIleGluHisIleuAlaIySGlnGlyAsnAArgPhe 60  
 Db 791 TCTCTCACCATATAGTGGTGGGAAAGCATATGACCATTTGGCCAAACAGGAATAGATT 860  
 QY 61 HisPheAspIleIyProProLeuHisIleAlaIySAsnCyseAspPheSerPheThrGly 80  
 Db 861 CATTTTGCATCAATCAATCTCTGTCATCATGCTTAATAATTTGATTTTCTTTTACATGGA 910  
 QY 81 LeuGlnHisValThrAspIyIleIleMetIySValGluLeuGluGlyIleGluIyS 100  
 Db 911 CTTCAACACGTTATCGATTAATAATATGAAACAGGAAACAGAGAGGATTTGAGAG 970  
 QY 101 GlyGlnIleLeuSerSerIleAlaAspIleAlaIleValGlnHisIleThrMetAlaCys 120  
 Db 971 GGGCAAAATCTGCTTTCAGACAGACATTCGTCACAGTACAGCACCAATGCGATGT 1030  
 QY 121 HisLeuValIyAsArgThrHisArgAlaIleLeuPheCysIyGlnArgAspLeuLeuPro 140  
 Db 1031 CATCTTGCAAAAGAACACATCGGCTATTCCTGTTTGTAAAGACAGAGACTTGTTACCT 1090  
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyIyValAlaIleSerAsnPheTyrIleArgArg 160  
 Db 1091 CAAATAATATGACGTATCGTGTGCTGATCTGCTGTCGCAAGTAACTTATATATCCACAGA 1150  
 QY 161 AAlaLeuGluIleuThrAsnAlaThrGlnCysThrLeuLeuCyseProProArgLeu 180  
 Db 1151 GCTCTGGAATTTTAAACAAGCAACAGTGCCTTTGTGTGCTCTCCAGACTA 1210  
 QY 181 CysThrAspAnglyIleMetIleAlaTPanglyIleGluArgLeuArgAlaGlyLeu 200  
 Db 1211 TGCACGTATATATGCAATATATATGCAATGGAATGATTTGAAGAAGCTACGTCTGGCTTG 1270  
 QY 201 GAlIleLeuHisAspIleGluGlyIleArgTyrGluProLysCyseProLeuGlyValAsp 220  
 Db 1271 GGCATTTTACATGACATAGAAAGCAATCCGCTATGAACCAAAATGCTCTTGGAGTAAAC 1330  
 QY 221 ILeSerIySGluValGlyGluAlaSerIleIySValProGlnLeuIyMetGlyIle 239

Db	1331	ATATCAAAAGAGTTGAGAGAGCTTCATTAAGCTACACAAATTAAATGAGATA	1387
RESULT 7			
ID	ACA60887	standard; cDNA; 1820 BP.	
XX	ACA60887;		
XX	08-JUL-2003	(first entry)	
DE	Human cDNA 28472 encoding a glycoprotease.		
XX			
KW	Human; ss; gene; cancer; aberrant cellular proliferation;		
KW	differentiation; immune disorders; heart disorder; brain disorder;		
KW	cardiovascular disorder; endothelial cell disorder; pain disorder;		
KW	haematopoietic disorder; blood vessel disorder; metabolic disorder;		
KW	liver disorder; platelet disorder; glycoprotease.		
OS	Homo sapiens.		
FX			
FX	Key	Location/Qualifiers	
FX	CDS	146..1390	
FT		/tag=a	
FT		/product="glycoprotease"	
FT		/note="This CDS is specifically claimed in claim 1"	
XX	US2003009017-A1.		
XX	09-JAN-2003.		
XX	08-NOV-2001; 2001US-00012140.		
XX	08-NOV-2000; 2000US-0246768P.		
XX	08-NOV-2000; 2000US-0246772P.		
XX	15-NOV-2000; 2000US-0249185P.		
PA	(LEIB/) LEIBY K R.		
PA	(KAPE/) KAPPELLER-LIBERMANN R.		
PA	(GLUC/) GLUCKSMANN M A.		
PI	Leiby KR, Kapeller-Libermann R, Glucksmann MA;		
XX	WPI; 2003-428888/40.		
DR	P-PSDB; ABU09569.		
XX			
PT	New isolated 38650, 28472, 5495, 65507, 81588 and 14354 nucleic acid		
PT	molecules, useful for diagnosing, treating cancer, pain, or immune,		
PT	heart, endothelial cell, hematopoietic, blood vessel, brain, metabolic		
PT	and liver disorders.		
PS	Claim 2; Fig 8; 90pp; English.		
XX			
CC	The invention relates to an isolated 38650 (encoding adenosine		
CC	deaminase), 28472 (encoding a glycoprotease), 5495 (encoding a 7		
CC	transmembrane receptor), 65507, 81588 or 14354 nucleic acid molecule (or		
CC	a sequence which is at least 60% identical to the six nucleic acids or		
CC	their open reading frames, fragments of at least 15 nucleotides,		
CC	naturally occurring variants, or a DNA insert of the plasmid deposited		
CC	with the American Type Culture Collection as Accession No. not defined in		
CC	the specification, which encodes the amino acid sequence). Also included		
CC	are a host cell containing the nucleic acids (used to produce the		
CC	proteins), the encoded proteins, an antibody that selectively binds to		
CC	the polypeptide, and identifying a compound that binds to/modulates the		
CC	activity of the polypeptide. The nucleic acid molecules, polypeptides and/or		
CC	methods are useful for diagnosing, treating cancer, aberrant cellular		
CC	proliferation and/or differentiation, immune disorders, heart disorders,		
CC	cardiovascular disorders including endothelial cell disorders,		
CC	haematopoietic disorders, blood vessel disorders, brain disorders, pain		
CC	and metabolic disorders, liver disorders and platelet disorders (many		
CC	examples of these disorders are given in the specification). The present		
CC	sequence is the Human cDNA 28472 encoding a glycoprotease		
XX			

SQ	Sequence	1820 BP; 543 A; 365 C; 393 G; 518 T; 0 U; 1 Other;
 Alignment Scores:		
Pred. No.:	8,22e-128	Length: 1820
Score:	1203.00	Matches: 232
Percent Similarity:	98.33%	Conservative: 3
Best Local Similarity:	97.07%	Mismatches: 4
Query Match:	97.02%	Indels: 0
DB:	8	Gaps: 0
 US-10-649-273-2_COPY_176_414 (1-239) x ACA60887 (1-1820)		
OY	1	LeuLeuAlaleuValGInGlyValSerAspPheLeuLeuGlyIysSerLeuAspIle 20
Dd	671	CTGTGGCATTTAGTTCACAAAGATTTCAGATTTTCCTTGCGAAGAATCTTGACATA 730
OY	21	AlAProGIAspMctLeuAspIysValAlArgrRfouSerLeuIlelVshIAProGLu 40
Dd	731	GCACCAGGTGCACAGCTTGACCAAGGTGCACAAAGAACCTTCTTTAAATPAACATCCAG 790
OY	41	CysSerThrMetSerGIgIyLyValAlIegluHisLeuAlalysGInGIyAsnArgPhe 60
Dd	791	TGCTCCACCAAGATGGTGGGAAGCATTGAACAATTTGGCCAACAAGGAANAGATT 850
OY	61	HIsPheAspIlIelVSPProFouLeuHIShIALalysAnCyAspPheSerPheThrGIy 80
Dd	851	CATTTTGCACATCAACACTCCCTTCATCACAGCTAAAATTGTGATTTTTCTTTTACGGA 910
OY	81	LeuGlnHisValIThrAspIylsIleIemetLysSgluIysGluGluGlyIleGluLys 100
Dd	911	CTTCAACACGTTACTGATTAATAATATGAAAAACGGAACCAAGGAAGTATTGAGAG 970
OY	101	GIyGlnIleLeuSerSerAlaAlaAspIlIeAlaIatThrValGlnHIStrMetAlaCys 120
Dd	971	GGCGAATCTCGTCTTGACGACGACAGACTTCCTGCCACATGACGACACAAATGGCATGT 1030
OY	121	HisLeuValIylsArgThrHISargAlaIleLeuPheCysIysGlnArgAspLeuLeuPro 140
Dd	1031	CATCTTGGAAAAAGAACACATCGGGCTATTCTGTTTTGTAAAGACAGAGACTTGTA 1090
OY	141	GlnAsnAsnAlaValIleuValAlaSerGIyGIyAlaAlaserAsnPhetYrlIeArgArg 160
Dd	1091	CAAAATATATGACAGTACGTGTGCATCGGGGTGTCCCAAGTAACTTCTATATCCGAGA 1150
OY	161	AlAleuGlnIleLeuThrAsnAlaThrGlnCysThrIleuLeuCysProProArgIeu 180
Dd	1151	GCTCGGAATTTTAAACAAGCAACAGTGCACITGTGTGTCTCTCCACAGACTA 1210
OY	181	CysThrAspAsnGIylemetIleAlaTrpAsnGIyleGluArgLeuArgAlaGIyLeu 200
Dd	1211	TGCACGTATATATGSCATTATGATTGCATGGAAATGTATTAAGAAGCTACGTGCGTTG 1270
OY	201	GIyIleLeuHisAspIlIeGIyGIyIleArgTYrGluProLysCysProLeuGIyValAsp 220
Dd	1271	GGCATTTTACATGACATAGAAAGCATCCGCTATCAACAACAAAAGTCTCTTGGAGTAGAC 1330
OY	221	IleSerIysGIyValGIyGluAlaSerIleLysValProGlnIleuIysMetGluIle 239
Dd	1331	ATATCAAAGAAGATGGAGAAGCTTCCATTAAGTACCAATTTAAAAATGAGAGATA 1387
 RESULT 8 ABSS7020 ID ABSS7020 standard; cDNA; 1821 BP.		
 XX AC ABSS7020; XX DT 30-JAN-2003 (first entry) XX DR cDNA encoding novel human glycoprotease 26472. XX KW Cancer; aberrant cell proliferation; aberrant cell differentiation; XX breast cancer; ovarian cancer; prostate cancer; colon cancer; XX lung cancer; immune disorder; heart disorder; cardiovascular disorder;		

KW	endothelial disorder; hematopoietic disorder; blood vessel disorder;
KV	brain disorder; pain; metabolic disorder; liver disorder; diabetes;
KM	platelet disorder; carcinoma; sarcoma; leukemia; Hodgkin's disease;
KN	autoimmune disorder; hypertension; atherosclerosis; heart failure;
KP	myocardial infarction; ischemic heart disease; Crohn's disease;
KQ	Grave's disease; Kawasaki syndrome; Raynaud's disease; aneurysm;
KR	cerebral ischemia; peripheral neuropathy; Alzheimer's disease;
KS	Parkinson's disease; anorexia nervosa; cachexia; glycoprotease; gene.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	147..1391
FT	/tag=a
FT	/product="Glycoprotease 28472"
XX	/note="Specifically claimed in claim 1"
PN	WO200277233-A2.
PD	
PF	03-OCT-2002.
XX	
PR	08-NOV-2001; 2001WO-US04672A.
PR	08-NOV-2000; 2000US-024676B.
PR	08-NOV-2000; 2000US-024677Z.
PR	15-NOV-2000; 2000US-024918SP.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
PI	
P1	Lelby KR, Kapeller-Libermann R, Glucksmann M;
XX	
DR	WPI: 2003-029938/02.
DR	P-FSDB; ABG71162.
XX	
PT	New adenosine deaminase, glycoprotease and seven transmembrane domain
PT	nucleic acids and polypeptides, designated 38650, 28472, 5495, 65507,
PT	81588 and 14354, useful for treating e.g. leukemias, Hodgkin's disease or
PT	hypertension.
XX	
PS	Claim 2; Fig 8A-B; 178pp; English.
XX	
CC	The invention describes isolated 38650, 28472, 5495, 65507, 81588 and
CC	14354 nucleic acid molecules (I) and their encoded polypeptides (II). The
CC	38650 nucleic acid molecule comprises a sequence encoding adenosine
CC	deaminase. The 28472 nucleic acid molecule comprises a sequence encoding
CC	a human glycoprotease. The 5495, 65507, 81588 and 14354 comprise
CC	sequences that encode a human seven transmembrane domain (7TM). The
CC	38650, 28472, 5495, 65507, 81588 or 14354 nucleic acid and polypeptide
CC	sequences are useful for diagnosing, preventing or treating a subject
CC	with or at risk of developing a disorder, e.g. cancer or aberrant
CC	cellular proliferation and/or differentiation (e.g. breast, ovarian,
CC	prostate, colon or lung cancer), immune disorders, heart disorders,
CC	cardiovascular disorders, endothelial disorders, hematopoietic disorders,
CC	blood vessel disorders, brain disorders, pain and metabolic disorders,
CC	liver disorders or platelet disorders. These disorders include carcinoma,
CC	sarcoma, leukemia, Hodgkin's disease, autoimmune disorders,
CC	hypertension, atherosclerosis, heart failure, myocardial infarction,
CC	ischemic heart disease, Crohn's disease, Grave's disease, Kawasaki
CC	syndrome, Reynaud's disease, aneurysm, cerebral ischemia, peripheral
CC	neuropathy, Alzheimer's disease, Parkinson's disease, anorexia nervosa,
CC	cachexia or diabetes. This sequence encodes the novel human glycoprotease
CC	28472
XX	
SQ	Sequence 1821 BP; 543 A; 365 C; 394 G; 518 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	8,22e-128	length:	1821
Scores:	1203.00	Matches:	232
Percent Similarity:	98.33%	Conservative:	3
Best Local Similarity:	97.07%	Mismatches:	4
Query Match:	97.02%	Indels:	0
GB:	10	Gaps:	0

	US-10-649-273-2_COPY_176_414 (1-239) x ABSS7020 (1-1821)
OY	1 LeuLeuAlaIeuValGInGlyValSerAspPheLeuLeuGlyIysSerLeuAspIle 20
Db	672 CTGTGGCATTAAGTCAAGAGTTCCAGATTTCCTGCTTCTTGGAAGCTTTTGACATA 731
OY	21 AlaProGlyAspMetLeuAspIlyValAlaArgArgIeuSerLeuIleIysHisProGlu 40
Db	732 GCACACAGGTGCATGCTTGCACAGGTGGCAAGAAAGACTTCTTTAAATPAAACATCCAGAG 791
OY	41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaIysGInGlyAsnArgPhe 60
Db	792 TGCTCCACCATGATGGTGGGAAAGCCATAGAACATTGGCCAAACAGAAATAGATT 851
OY	61 HisPheAspIleIysProProLeuHisHisAlaIysAsnCysAspPheSerPheThrGly 80
Db	852 CATTTTGACATCAAACTCCCTTCATCATCTCTAAATATGTGATTTTCTTTTACCTGGA 911
OY	81 LeuGlnHisValIThrAspIlyIleIleMetIlyIysGluIysGluGluGlyIleGluIys 100
Db	912 CTTCAACACGTTACTGATTAATAATATGAAAAACGGAACAAAGGAAGATTTGAAAGAG 971
OY	101 GlyGlnIleIeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisSerMetAlaCys 120
Db	972 GGGCAATCTCTGTCTTGACGACGACGACATTCCTCCACAGACGACACATGCGCATGT 1031
OY	121 HisLeuValIlySarGlyHisSarGAlaIleLeuPheCysIlyGlnArgAspLeuLeuPro 140
Db	1032 CATCTTGGAATAAGAACACATCGGGCTATTCTGTTTGTGAAGCAGAGACTTGTTACT 1091
OY	141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAsnAsnPheIyrIleArgArg 160
Db	1092 CAAATATATGACAGTACGTGTGATCTGTGTGTGTCCAAAGTAACTTCTATATCCGAGA 1151
OY	161 AlaIeuGluIleLeuThrAsnAlaThrGlnCysThrIleLeuCysProProArgLeu 180
Db	1152 GCTCTGAAATTTTAAACAAAGCAACAGTGCCTTTGTTGTCTCTCTCCAGACTA 1211
OY	181 CysThrAspAsnGlyIleMetIleAlaTPAsnGlyIleGluArgLeuArgAlaGlyLeu 200
Db	1212 TGCACTGATTAATGGCATTTATGATTCATGGAATGTAATTGAAACATACGTGGCTTG 1271
OY	201 GlyIleIeuHisAspIleGluGlyIleArgArgIleGluProIysCysProLeuGlyValAsp 220
Db	1272 GGCATTTTACATGACATAGAAAGGATCGCTATACCAAAATGTCTCTTGAGTAGAC 1331
OY	221 IleSerIysGluValGlyGlnAlaSerIleIysValProGlnLeuIysMetGluIle 239
Db	1332 ATATCCAAAGAAAGTTGGAGAAAGCTTCATTAATAAGTACCAATTAATAATGAGATA 1388
RESULT 9	
ADA52832	
ID	ADA52832 standard; cDNA; 2208 BP.
AC	ADA52832;
XX	
XX	20-NOV-2003 (first entry)
XX	
DE	Human coding sequence, SEQ ID 400.
XX	
KW	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW	Gene therapy; human; secretory protein; membrane proteins; cancer;
KW	Inflammatory disease; osteoporosis; neurological disease; gene; ss.
OS	Homo sapiens.
XX	
XX	EP1293569-A2.
XX	
XX	19-MAR-2003.
XX	
XX	21-MAR-2002; 2002EP-00006586.
XX	
XX	14-SEP-2001; 2001JP-00328381.
PR	



PR 24-JAN-2002; 2002US-0350435P.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isega T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
 XX  
 DR WPI: 2003-395539/38.  
 DR P-PSDB; ADA54471.  
 XX  
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.  
 XX  
 PS Claim 1; SEQ ID NO 400; 205bp; English.  
 XX  
 CC The present invention relates to novel human secretory or membrane  
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.  
 XX  
 SQ Sequence 2208 BP; 660 A; 454 C; 451 G; 643 T; 0 U; 0 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 3.88e-111 Length: 2208  
 Score: 1059.00 Matches: 211  
 Percent Similarity: 88.28% Conservative: 0  
 Best Local Similarity: 88.28% Mismatches: 4  
 Query Match: 85.40% Indels: 24  
 DB: 10 Gaps: 1  
 US-10-649-273-2\_COPY\_176\_414 (1-239) x ADA52832 (1-2208)  
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLeuSerLeuAspLeu 20  
 Db CTGTTGGCATTAGTTCAGAGAGATTTCAGATTTCCTTGGAAAGCTTTTGACATTA 928  
 QY 21 AlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleValHisProGlu 40  
 Db GCACCAAGGTGACATGCTTGACAGAGGTGGCAAGAAAGCTTCTTATATAAATCAGAG 988  
 QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaValGlnGlyAsnArgPhe 60  
 Db TGCCTCCACCAATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAGGAAATGATT 1048  
 QY 61 HisPheAspIleLeuProProLeuHisHisAlaValAsnCysAspPheSerPheThrGly 80  
 Db CATTTTGCATCAAACTCCCTTGATCATGCTAAATAATGATTTCTTTTAACTGGA 1108  
 QY 81 LeuGlnHisValThrAspLeuIleIleMetCysLeuGluValGluGlyIleGluLeu 100  
 Db CTTCAACACGTTACTGATTAATAATATATGAAAAAGAAAAAGAGATTTGAGAG 1168  
 QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
 Db GGGCAAAATCCGTCTTGACAGACAGACATGTCTGCCACAGTACAGACACAAATGCGAT 1228  
 QY 121 HisLeuValIysArgThrHisArgAlaIleLeuPheCysLeuGlnArgAspLeuPro 140  
 Db CATCTTGTGAAGAAACACATCGGCTATCTGTTTGTAAAGACAGAGACTGTACTCT 1288  
 QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160  
 Db CAAAATTAATGACAGATCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1348  
 QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180  
 Db GCTCTGGAATTTTAAACAAACAGCAACAGTGCATTTGTGTGCTCTCTCCAGACTA 1408  
 QY 181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgAlaGlyLeu 200

Db 1409 TGCACGTATTAATGACATTAATGATTCGA----- 1435  
 QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProIleCysProLeuGlyValAsp 220  
 Db 1436 -----TGAATGCTCTCTTGGAGTAGAC 1456  
 QY 221 IleSerTyrGlyValGlyGlnAlaSerIleValValProGlnLeuIleCysMetGluIle 239  
 Db 1457 ATATCAAAAGAACTTGGAGAGAGCTTCCATTAAGTACCAATTTAAATAATGAGATA 1513  
 RESULT 10  
 ADQ24627  
 ID ADQ24627 standard; DNA; 2890 BP.  
 XX  
 AC ADQ24627;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7447.  
 XX  
 KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;  
 KW ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004046938-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 26-NOV-2003; 2003WO-US038193.  
 XX  
 PR 26-NOV-2002; 2002US-0429739P.  
 XX  
 PA (PROT-) PROTEIN DESIGN LABS INC.  
 XX  
 PI Aziz N, Ginsburg WM, Zlotnick A;  
 XX  
 DR WPI: 2004-441208/41.  
 XX  
 PT Early detection of soft tissue sarcoma comprises determining expression  
 PT of a gene in a first soft tissue sample and a normal soft tissue sample  
 PT and comparing the gene expression, also useful in treating soft tissue  
 PT sarcoma.  
 XX  
 PS Example 2; SEQ ID NO 7447; 210bp; English.  
 XX  
 CC The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cytostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC DNA of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.  
 XX  
 SQ Sequence 2890 BP; 869 A; 609 C; 611 G; 789 T; 0 U; 12 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 5.67e-111 Length: 2890  
 Score: 1059.00 Matches: 211  
 Percent Similarity: 88.28% Conservative: 0  
 Best Local Similarity: 88.28% Mismatches: 4  
 Query Match: 85.40% Indels: 24  
 DB: 12 Gaps: 1  
 US-10-649-273-2\_COPY\_176\_414 (1-239) x ADQ24627 (1-2890)



```

QY      1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB      1526 CTGTTGGCATTAGTTCAGAGGTTTCAGATTTTCCTTCTTGGAAAGCTTTGGACATTA 1585
QY      21 AlaProGlyAspMetLeuAspLysValAlaArgGluLeuSerLeuIleLysHisProGlu 40
DB      1586 GCACACAGGTGACATGCTTGACAGAGTGCAGAGAACATCTCTTATATAAACATCCAGAG 1645
QY      41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaValGlnGlyAsnArgPhe 60
DB      1646 TGCTTCACCATGAGTGTGGAGAGCCATAGAACCTTTGGCCAAACAGAAATATGATTT 1705
QY      61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB      1706 CATTGGACATCAAACTCCCTTGATCATGTGTAATAAATTTGGATTTCTTTTACTGGA 1765
QY      81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
DB      1766 CTTCAACACGTTACTGATTAATAATATATGAAAAAGAAAAAGAGATTTGAGAAAG 1825
QY      101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetLacCys 120
DB      1826 GGGCAAAATCCCTGCTTCAGCAGCAGACATGCTGCACAGATACACACCAATGGCATGT 1885
QY      121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB      1886 CATCTTGGAAAAAGAACACATCCGGCTATTCCTGTTTGTAAACAGAGACTGTTACTT 1945
QY      141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaAsnAsnPheThrIleArgArg 160
DB      1946 CAAATATATGACAGTACTGTTGATCTGTGTGTGTCGCAAGTACTTCTGATCCGACAG 2005
QY      161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrIleLeuCysProProProArgLeu 180
DB      2006 GCTCTGGAATTTTAAACAACGACACACAGTCTTGTGTGTCTCTCCACAGACTA 2065
QY      181 CysThrAspAsnGlyIleMetIleAlaThrAsnGlyIleGluArgLeuArgIleGluLeu 200
DB      2066 TGCACTGATATGTCATTAATGATTCGA----- 2092
QY      201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
DB      2093 -----TGATGTCTCTTGGAGTAGAC 2113
QY      221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB      2114 ATATCAAAAGAGTGGAGAGCTTCCATAAAGTACCAATTAATAAATGAGAGATA 2170

```

RESULT 11  
ADE31345/c  
ADE31345 standard; DNA; 3358 BP.

```

ID      ADE31345 standard; DNA; 3358 BP.
XX
AC      ADE31345;
XX
DT      29-JAN-2004 (first entry)
XX
DE      Human diagnostic and therapeutic polynucleotide (dithp), SEQ ID NO 100.
XX
XX      diagnostic and therapeutic polynucleotide; dithp; antiarteriosclerotic;
XX      antiinflammatory; cerebroprotective; antidiabetic; antidiabetic;
XX      immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquillizer;
XX      osteopathic; antiarthritic; antirheumatic; cytostatic; hepatotropic;
XX      virucide; haemostatic; anti-HIV; antithyroid; thyromimetic;
XX      dermatologic; antibacterial; fungicide; antiparasitic; anticonvulsant;
XX      thrombolytic; anticoagulant; anorectic; vasodilator; antidiabetic;
XX      gene therapy; protein replacement therapy; human; gene; ds.
XX
OS      Homo sapiens.
XX
XX      WO2003062376-A2.
XX
PD      31-JUL-2003.
XX

```

```

PR      13-JAN-2003; 2003WO-US001096.
XX
PR      16-JAN-2002; 2002US-0349384P.
PR      17-JAN-2002; 2002US-0349413P.
PR      17-JAN-2002; 2002US-0349946P.
XX
XX      (INCY-) INCYTE GENOMICS INC.
PI      Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JH,
PI      Yu JY, Tudson O, Yap PB, Amesey SR, Dam TC, Liu TP, Gerstin EH,
PI      Perilla CH, Lewis SA, Chen A, Marwaha R, Lan RX, Uraehka ME,
PI      Kristnam SR, Kolluru V, Panesar IS;
XX
DR      WPI: 2003-636732/60.
XX
DR      P-PSDB; ADE31156.
XX
XX
PT      New human diagnostic and therapeutic polynucleotides and polypeptides,
PT      useful for diagnosing, treating or preventing e.g. leukemia, brain
PT      cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke
PT      or Alzheimer's.
XX
PS      Claim 1; SEQ ID NO 100; 634bp; English.
XX
XX
CC      The invention relates to a novel isolated human diagnostic and
CC      therapeutic polynucleotide (designated dithp). The novel dithp
CC      polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798
CC      base pairs fully defined in the specification; a polynucleotide
CC      comprising a naturally occurring polynucleotide sequence at least 90%
CC      identical to the dithp polynucleotide; a polynucleotide complementary to
CC      the dithp polynucleotide or its polynucleotide which is at least 90%
CC      identical; or an RNA equivalent of any of the polynucleotides mentioned
CC      above. The dithp polynucleotides have the following activities:
CC      antiarteriosclerotic, antiinflammatory, cerebroprotective, antidiabetic,
CC      antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,
CC      tranquilizer, osteopathic, antiarthritic, antirheumatic, cytostatic,
CC      hepatotropic, virucide, haemostatic, anti-HIV, antithyroid, thyromimetic,
CC      dermatologic, antibacterial, fungicide, antiparasitic, anticonvulsant,
CC      thrombolytic, anticoagulant, anorectic, vasodilator, and antidiabetic. The
CC      novel dithp polynucleotides polypeptide can be used in gene therapy and
CC      protein replacement therapy. The dithp polynucleotides or dithp
CC      polypeptides are useful for diagnosing, preventing or treating diseases
CC      associated with the expression of human molecules. In particular, these
CC      diseases include cancers (e.g. adenocarcinoma, leukaemia, melanoma, brain
CC      cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung
CC      cancer) or other cell proliferative disorders (e.g. arteriosclerosis,
CC      atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary
CC      thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,
CC      Addison's disease, thyroiditis, Crohn's disease, Graves' disease,
CC      Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
CC      arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.
CC      viral, bacterial, fungal or parasitic infection), developmental disorders
CC      (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.
CC      thrombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic
CC      disorders (e.g. hypercholesterolaemia, hypoglycaemia, diabetes,
CC      hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic
CC      cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,
CC      Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,
CC      anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),
CC      transport disorders (e.g. akinesia or multidrug resistance), or
CC      connective tissue disorders (e.g. Paget's disease or rickets). This
CC      polynucleotide sequence represents one of the human dithp DNA sequences
CC      of the invention.
XX
XX
SQ      Sequence 3358 BP; 1105 A; 577 C; 601 G; 1075 T; 0 U; 0 Other;

```

## Alignment Scores:

```

Pred. No.: 1,4e-103
Score: 995.50
Percent Similarity: 87.828
Best Local Similarity: 87.828
Query Match: 80.288
DB: 10

```

Length: 3358

Matches: 209

Conservative: 0

Mismatch: 2

Indels: 29

Gaps: 1

```

US-10-649-273-2_copy_176_414 (1-239) x ADE31345 (1-3358)
QY 29 ValAlaArgArgLeuSerLeuIleValHisProGluCysSerThrMetSerGly-GlyIle 48
DB 2104 GTGGCAGAGAGACTTTCTTTAATATTAACATCAAGAGCTCCACCATGAGGGGGAA 2045
QY 48 salaiiegluHisleuValaysgingIlyAsnArgPheHisPheAspIleLeuProProle 68
DB 2044 AGCCATATGAACATTGGCCAAACAGAGAAATGATTTTCATTTTGACATCAAACTCCCTT 1985
QY 68 uHisHisAlaIleValAsnCyAspPheSerPheThrGlyLeuGlnHisValThrAspIle 88
DB 1984 GCATCATGCTGCTAAAATATGATTTCTTTTACTGACCTCAACACCTTACTGATTAAT 1925
QY 88 eileMetIleValysGluIlyysGluIly----- 96
DB 1924 AATTAATGAAAAAGAAAAAGAGAGGATATTTCTAATTAAGTAAAGTTGAACAGATAAA 1865
QY 97 -----GlyIlyegluIlyysGluIly 103
DB 1864 TATTCCTGATGTGCTTAAAAAAGCTGCTCATTTCTGACAGGTATTCAGAGGGGAAA 1805
QY 103 leIeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCysHisIleuV 123
DB 1804 TCTGCTCTTACAGCAGACATTTGCTGCAAGTACAGACACAAATGGCATGTCTATCTTG 1745
QY 123 allYsArgThrHisArgAlaIleleuPheCysIlyysGluIlyAspIleuProGlnAsn 143
DB 1744 TGAAGAGAACACATCGGGCTAATCTGTTTGTAGCAGAGAGACTGTTTACTCAAAATA 1685
QY 143 snAlaValIleuValAlaSerGlyIlyValAlaSerAsnPheTyrIleArgArgAlaIleu 163
DB 1684 ATCCAGTACTGTTGCTCTGT-GGTGTCCGAAATTAATTTAATCCGACAGCTCTGG 1626
QY 163 IuIleuThrAsnAlaThrGlnCysThrIleuLeuCysProProArgLeuCysThrA 183
DB 1625 AATTTTAAAC-AACGCAACACAGTCACTTGTGTGTCTCCCTCCACGATAGCACTG 1567
QY 183 sPAsnGlyIleMetIleAlaIleTTPAsnGlyIlyegluIlyysGluIlyysGlyIlyel 203
DB 1566 ATATATGCAATTATGATTCATGATGAAATGATTAAGAAAGCTACGTCGTGGCATTT 1507
QY 203 euHisAspIleGluIlyIleArgTyrGluProIlyCysProIleuGlyValAspIleSerL 223
DB 1506 TACATGACATAGAGAGCATCCGCTATGAAACCAAAATGTCTCTTGAGTAAACATATCAA 1447
QY 223 ysgIuValIlyGlyIlyAlaSerIleIysValProGlnIleuIlyMetGluIle 239
DB 1446 AAGAAGTTGAGAGAAAGCTTCATTAATAAGTACCAATTAATAATGAGATA 1397
RESULT 12
ABQ75508
ID ABQ75508 standard; DNA; 1572 BP.
XX
AC ABQ75508;
XX
DT 07-NOV-2002 (first entry)
XX
DE Murine sialoglycoprotease-like gene sequence SEQ ID NO:7.
XX
KW Murine; mouse; protease; calcium activated neutral protease type 5;
KW CAPN5; trypsinase 4; sialoglycoprotease; enzyme; genetic disease;
KW neurological; neuropsychological; psychotic illness; transgenic animal;
KW gene; ds.
XX
OS Mus musculus.
XX
PN WO200245491-A2.
XX
PD 13-JUN-2002.
XX
PP 05-DEC-2001; 2001MO-US046405.
XX

```

```

PR 06-DEC-2000; 2000US-0251803P.
PR 06-DEC-2000; 2000US-0251820P.
PR 13-DEC-2000; 2000US-0255971P.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Allen KD, Leviten MW;
XX
DR WPI: 2002-657389/70.
XX
PT Novel transgenic animal, comprising a disruption in protease target gene,
PT is useful for identifying agents that ameliorates a phenotype associated
PT with a disruption in a protease target gene.
XX
PS Example 3; Fig 7; 62bp; English.
XX
CC The present invention describes a non-human transgenic animal (I)
CC comprising a disruption in a protease target gene (PG) selected from
CC calcium activated neutral protease type 5 (CAPN5) gene, trypsinase 4 gene
CC and sialoglycoprotease-like gene. Also described is a targeting construct
CC (II), comprising a first polynucleotide sequence homologous to at least a
CC first portion of PG, a second polynucleotide sequence homologous to at
CC least a second portion of PG and a selectable marker. (II) is useful for
CC producing a transgenic mouse comprising a disruption in a protease target
CC gene, by introducing (II) into a cell, introducing the cell into a
CC blastocyst, implanting the resulting blastocyst into a pseudopregnant
CC mouse, where the pseudopregnant mouse gives birth to a chimeric mouse,
CC and breeding the chimeric mouse to produce the transgenic mouse. (I) is
CC useful for identifying an agent that modulates the expression or function
CC of a protease target gene, by administering an agent to (I) and
CC determining whether the expression or function of the disrupted protease
CC target gene in (I) is modulated. (I) is also useful for testing the
CC efficacy of proposed genetic and pharmacological therapies for human
CC genetic diseases, such as neurological, neuropsychological or psychotic
CC illness. The present sequence represents murine sialoglycoprotease-like
CC gene sequence, which is used in an example from the present invention
XX
SQ Sequence 1572 BP; 459 A; 337 C; 340 G; 429 T; 0 U; 7 Other;
Alignment Scores:
Pred. No.: 1.28e-89 Length: 1572
Score: 870.00 Matches: 184
Percent Similarity: 82.77% Conservative: 13
Best Local Similarity: 77.31% Mismatches: 37
Query Match: 70.16% Indels: 5
DB: Gaps: 2
US-10-649-273-2_copy_176_414 (1-239) x ABQ75508 (1-1572)
QY 1 LeuLeuAlaIleuValGlnGlyValSerAspPheLeuLeuGlyIlySerLeuAspIle 20
DB 528 CTGTTGGKRTTAGTTCAGAGGTGTTCCTCAATTCCTGCTCCCTGGAAAGCTTTGACAAA 587
QY 21 AlaProGlyAspMetLeuAspIlyValAlaArgArgLeuSerLeuIleHisIspGlu 40
DB 588 GCACCGAGCAATGCTTGACAGAGTGGCAAAAGACTTTCTTAATCAACATCCAGAA 647
QY 41 CysSerThrMetSerGlyIlyValAlaIleGluHisLeuAlaIlysgingIlyAsnArgPhe 60
DB 648 TCTTTCAATGAGAGTGGTGAAGAACTATAGAACAGTTGCCAAAGACGAAATGATTC 707
QY 61 HisPheAspIleIysProProIleuHisAlaIlyAsnCyAspPheSerPheThrGly 80
DB 708 CATTTTACTATCAACACCATAGTACAGATCTAAGAAATTCGATTTTCTTCAACGGGA 767
QY 81 LeuGlnHisValThrAspIleIleIleMetIlyysGluIlyysGluIlyegluIly 100
DB 768 CTTCACATATTAATGATTAAGCTAATTAACACAGAGAAAAAGAAAGCAATTAAGAAG 827
QY 101 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 828 GGGCAATCTGTCTATGCTGACAGCATTTGCTGCGGTACAGATGCAACAGCGTGC 887

```

QY 121 HisLeuValIysArgThrHisArgAlaIleuPheCysGlnArgAspLeuPro 140  
DB 888 CACCTTGCAGAAAGAACACATCGCGTATTTCTGTTTCCAGACAGAAAAATTTGCTCT 947  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160  
DB 948 CCAGCTAACCCAGATTAAGTGTATCTGAGAGGTGTGCAAAATGCTTACATCCGAAAA 1007  
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrIleuLeuCysProProArgLeu 180  
DB 1008 GCATGTGAATATGTCGCAAAATGCAACGATGCAACGATGTCCTTCCACT-TCAGACTG 1066  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 1067 TGCACCTGACAAATGGCATATTCATTCATGAAATGGAAATGTAACGTCCTGCTT 1126  
QY 201 GlyIleLeu-HisAsp-IleGluGlyIleArgTyrGluProLysCysProLeuGlyVal- 219  
DB 1127 GGCCTTTTACCATGATGATAGACATTCGCTTATTAAACCAATGCTCCTTTGAGTGA 1186  
QY 220 --AspIleSerLysGluValGlyGluAlaSerIleLysValProGluLeu 235  
DB 1187 GGCATTATCCCGAAGAGTTGGCAGA--ACCTTGCCCATTTAAAAAGTTA 1233  
RESULT 13  
ID AAS84622 standard; cDNA; 2734 BP.  
XX AAS84622;  
AC AAS84622;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #20426.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001MO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Dmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-SDB; ABG20435.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 20426; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. M564197-M564564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2734 BP; 773 A; 545 C; 639 G; 763 T; 0 U; 14 Other;  
XX  
Alignment Scores:  
Pred. No.: 4,02e-58 Length: 2734  
Score: 599.00 Matches: 131  
Percent Similarity: 71.57% Conservative: 10  
Best Local Similarity: 66.50% Mismatches: 22  
Query Match: 48.31% Indels: 34  
DB: Gaps: 6  
US-10-649-273-2\_COPY\_176\_414 (1-239) x AAS84622 (1-2734)  
QY 43 ThrMetSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHisPhe 62  
DB 206 ACCCTGCTAGGGGGGGCGCT-----AGATTCAT----- 235  
QY 63 AspIleLysProProLeuHisAlaLys-----AsnCys----- 74  
DB 236 -----CCTATTCTCCGATGAAAGTATGACGCTGTATGAGAACGTCGACGA 283  
QY 75 -----AspPheSerPheThrGlyLeuGlnHisValThrAsp 86  
DB 284 TCTGCTTCTGGGAGAGCCCGAGAGCTTTACTCATGCGGAGAAAGTGGAGATGAG 343  
QY 87 LysIleIleMetLysGlyLysGluGlu-----GlyIleGluGlyGlnIleLeu 104  
DB 344 CGCTTACATGACAGACAGACAGACCGAGAGAGGGGGGATTTGAGAGGGGCAATCTCG 403  
QY 105 SerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCysHisLeuValLys 124  
DB 404 TCTTCAGACAGACATGCTGCTGCCACAGTACAGCACACAAATGGATGTCACTTGTGAAA 463  
QY 125 ArgThrHisArgAlaIleIleuPheCysLysGlnArgAspLeuProGlnAsnAla 144  
DB 464 AGAACACATCGGGCTATTCTGTTGTGTAGCAGAGAGACTGTGTACTCAAAATTAATGCA 523  
QY 145 ValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaLeuGluIle 164  
DB 524 GTCCTGCTTGCATCTGCTGTGTCTCCAGATTAATCTTATATCCGAGACTCTCGAAAT 583  
QY 165 LeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeuCysThrAspAsn 184  
DB 584 TTAACAAAGCAACACAGTGCCTTGTGTCTCTCCACAGACTATGACACTGATTAAT 643  
QY 185 GlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGlnHis 204  
DB 644 GGCATTATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 703  
QY 205 AspIleGluGlyIleArgTyrGluProLys-----CysProLeuGlyVal 219  
DB 704 GCATTAAGAGGATTCGCTATTAACCAAAATGATGTGCTCTTCCAGGGCTG 754  
RESULT 14  
ADL86725  
ID ADL86725 standard; DNA; 371 BP.  
XX ADL86725;  
AC ADL86725;  
XX  
DT 20-MAY-2004 (first entry)  
XX  
DE DNA up-regulated in murine multipotent progenitor cells Seqid 3118.

KW gene potential; multi-lineage; cell commitment; haematopoietic stem cell;  
KM HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;  
KM common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.  
XX  
OS Mus sp.  
XX  
XX WO2003093445-A2.  
XX  
XX 13-NOV-2003.  
XX  
XX 05-MAY-2003; 2003WO-US014114.  
XX  
XX 03-MAY-2002; 2002US-0377383P.  
XX  
XX (STOW-) STOWERS INST MEDICAL RES.  
XX  
XX L1 L;  
XX  
XX WPI; 2004-022656/02.  
XX  
XX  
XX Classifying an unknown multi-lineage affiliated gene comprises isolating  
PT expressed nucleic acid sequences from the discrete cell sub-populations.  
XX  
XX Claim 8; SEQ ID NO 3118; 123bp; English.  
XX  
XX  
CC This invention relates to a novel method for predicting gene potential by  
CC associating nucleic acid sequences of unknown function with particular  
CC sub-population profiles. Specifically, it refers to classifying an  
CC unknown multi-lineage affiliated gene by collecting hybridisation data to  
CC develop a gene expression map, in order to determine the discrete sub-  
CC population where it is expressed. The present invention describes methods  
CC for predicting the lineage commitment of genes associated with the self-  
CC renewing haematopoietic (blood) stem cells (HSCs), as well as the non-  
CC self renewing multipotent progenitors (MPPs), common lymphoid progenitors  
CC (CLPs) and common myeloid progenitors (CMPs), which are collectively  
CC referred to as bone marrow stem cells populations. As such, these methods  
CC can be used to identify associated multi-lineage affiliated genes and  
CC hence the underlying molecular mechanisms in physiological haematopoietic  
CC development. This polynucleotide sequence is DNA associated with a murine  
CC MPP sub population of cells of the invention.  
XX  
XX  
SQ Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;  
XX  
XX  
XX Alignment Scores:  
Pred. No.: 2.75e-44 Length: 371  
Score: 468.00 Matches: 92  
Percent Similarity: 78.86% Conservative: 5  
Best Local Similarity: 74.80% Mismatches: 26  
Query Match: 37.74% Indels: 0  
DB: Gaps: 0  
US-10-649-273-2\_COPY\_176\_414 (1-239) x ADL86725 (1-371)  
QY 109 Asptlleaiaathrrvalglnhsthmetlaacyshlsleuvallyaargthrhlsarg 128  
DB 2 GACATTGCTGCTGGCGGTAACAGACAGCAACAGCGTGCACCTTGCAGAAAGAAACACATGCC 61  
QY 129 AlalaleuPheCylysglnarAspleuLeuProglinaaanaAlaValleuVala1a 148  
DB 62 GCTATTCTGTTTCCAGACAGCAAGAAATTGCTCTCTCCAGCAACAGCGATATTACTTTGA 121  
QY 149 SerGlycylValAlaSerAenPheYrllleArGArgAlaLeuGluIlleuThraAna1a 168  
DB 122 TCTGGAGAGTGTGACAGTAACCTTGTACATCCGAAAGCATTTGGAATTTGCGCAATGCA 181  
QY 169 ThGlnCythrleuLeuCyProProbrArgLeuCystrThraaAngly1leuMet1le 188  
DB 182 ACCGAGTGCAGGTGTGTGTCTCCACTCCCAAGACTGTGCACATGCAATGCAATCATGATT 241  
QY 165 AlaftrPasnGlyllegluarGleuArgAlaGlyleuGlylleuHlsAsptllegluGly 208  
DB 242 GCATGGAATGGAATTGAANGATTACGTGCAGNCTTNGGCACTTTTACATGANTTAGAAGAC 301

QY 209 lIeargTyrgluProlyscYsProleuGlyValAsp1leSerlysgluVala1gylGluAla 228  
DB 302 ATCCNNTATGNNCCAAAATNTCTNTTGGAGTGAATNTCCAGAGAAGTTGCAAGACGT 361  
QY 229 Ser1leLys 231  
DB 362 GCCATMAAA 370  
RESULT 15  
ADL86726  
ID ADL86726 standard; DNA; 371 BP.  
XX  
XX ADL86726;  
XX  
XX 20-MAY-2004 (first entry)  
XX  
XX  
XX DNA up-regulated in murine multipotent progenitor cells SegID 3119.  
XX  
XX  
XX gene potential; multi-lineage; cell commitment; haematopoietic stem cell;  
KM HSC; multipotent progenitor; MPP; common lymphoid progenitor; CLP;  
KM common myeloid progenitor; CMP; bone marrow stem cell; mouse; murine; ds.  
XX  
XX Mus sp.  
XX  
XX WO2003093445-A2.  
XX  
XX 13-NOV-2003.  
XX  
XX 05-MAY-2003; 2003WO-US014114.  
XX  
XX 03-MAY-2002; 2002US-0377383P.  
XX  
XX (STOW-) STOWERS INST MEDICAL RES.  
XX  
XX L1 L;  
XX  
XX WPI; 2004-022656/02.  
XX  
XX  
XX Classifying an unknown multi-lineage affiliated gene comprises isolating  
PT expressed nucleic acid sequences from the discrete cell sub-populations.  
XX  
XX  
XX Claim 8; SEQ ID NO 3119; 123bp; English.  
XX  
XX  
XX This invention relates to a novel method for predicting gene potential by  
XX associating nucleic acid sequences of unknown function with particular  
XX sub-population profiles. Specifically, it refers to classifying an  
XX unknown multi-lineage affiliated gene by collecting hybridisation data to  
XX develop a gene expression map, in order to determine the discrete sub-  
XX population where it is expressed. The present invention describes methods  
XX for predicting the lineage commitment of genes associated with the self-  
XX renewing haematopoietic (blood) stem cells (HSCs), as well as the non-  
XX self renewing multipotent progenitors (MPPs), common lymphoid progenitors  
XX (CLPs) and common myeloid progenitors (CMPs), which are collectively  
XX referred to as bone marrow stem cells populations. As such, these methods  
XX can be used to identify associated multi-lineage affiliated genes and  
XX hence the underlying molecular mechanisms in physiological haematopoietic  
XX development. This polynucleotide sequence is DNA associated with a murine  
XX MPP sub population of cells of the invention.  
XX  
XX  
SQ Sequence 371 BP; 104 A; 75 C; 84 G; 92 T; 0 U; 16 Other;  
XX  
XX  
XX Alignment Scores:  
Pred. No.: 2.75e-44 Length: 371  
Score: 468.00 Matches: 92  
Percent Similarity: 78.86% Conservative: 5  
Best Local Similarity: 74.80% Mismatches: 26  
Query Match: 37.74% Indels: 0  
DB: Gaps: 0  
US-10-649-273-2\_COPY\_176\_414 (1-239) x ADL86726 (1-371)  
QY 109 Asptlleaiaathrrvalglnhsthmetlaacyshlsleuvallyaargthrhlsarg 128

```
Db      2  GACATTCGCTGCGGTACAGCATGCAACGCCGTCACCTTGCAGAAAAGAACACATCGC 61
Qy      129  AlalIleuPheCyIyGleGluArgAspIeuProGlnAsnAsnAlaValIleuValAla 148
Db      62  GCTATTCGTGTTTGCAAGCAGAGAAATTGCTCTCTCCAGCTAACGCAGTATTAGTTGTA 121
Qy      149  SerGlyGlyValAlaSerAsnPheTyrIleArgArgAlaIeuGluIleuThrAsnAla 168
Db      122  TCTGGAGGTGTGCAAGTACCTTGACATCCGAAAGCATTTGAAATTGTCCGCAATGCA 181
Qy      169  ThrGlnCysThrLeuLeuCysProProProArgLeuCyThrAspAsnGlyIleMetIle 188
Db      182  ACGCAGTGCACGTTGTGTGTCACCTCCAAAGACTGTGCACATGCAATGCAATGATT 241
Qy      189  AlatThrAsnGlyIleGluArgLeuArgAlaGlyLeuGlyIleuHisAspIleGluGly 208
Db      242  GCATGGAAATGGAATTGAAGATTACGTGCNGNCTTNGCNTTNTTACNTGANNTAGAAGAC 301
Qy      209  IleArgTyrGluProLysCysProIeuGlyValAspIleSerLysGlyValGlyIuAla 228
Db      302  ATCCNNTRATGNACCAAAATNTCNTTTTGAGTGAATTCATTCAGAGAACTTGCAAGAGCT 361
Qy      229  SerIleLys 231
Db      362  GCCATAAAA 370
```

Search completed: February 16, 2005, 13:59:01  
Job time : 468.139 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:19:37 ; Search time 142.101 Seconds  
(without alignments)  
2752.056 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 L1ALVQGVSPFL1LKGSLDI.....DISKVEGRASIKVPQLKMEI 239

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1202784 segs, 818138359 residues  
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODE=frame+;p2n.model -DEV=x1h  
-Q=/cgn2\_1/USPTO.spool/US10649273/runat\_14022005\_114704\_16425/app\_query.fasta\_1.1429  
-DB=Issued\_Patents\_NA -QPMT=fastcap -SUFFIX=rml -MINMATCH=0.1 -LOOPECL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCLLIG=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USBR=US10649273.qcgn\_1.1.292.qrunat\_14022005\_114704\_16425 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERRY -NEG SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	1416	4	US-09-774-528-177
2	1240	100.0	1526	4	US-10-067-443-23
3	1240	100.0	2197	4	US-10-067-443-1
4	1213.5	97.9	1387	4	US-10-067-443-21
5	725	58.5	14364	4	US-10-067-443-20
6	248.5	20.0	94750	4	US-09-596-002-38
7	247	19.9	1053	4	US-09-540-236-806
8	231	18.6	1059	4	US-09-252-991A-884
9	231	18.6	1205	4	US-09-252-991A-801
10	224	18.1	1830121	4	US-09-557-884-1
11	224	18.1	1830121	4	US-09-557-884-1
12	221.5	17.9	996	4	US-09-902-540-6612

C	13	221.5	17.9	2582	4	US-09-302-540-503	Sequence 503, App
	14	212	17.1	1074	4	US-09-543-681A-2341	Sequence 241, Ap
	15	205	16.5	1032	4	US-09-489-039A-2050	Sequence 2050, Ap
	16	200	16.1	1315	1	US-08-087-797-1	Sequence 1, Appli
	17	197	15.9	1092	1	US-09-107-532A-2955	Sequence 2955, Ap
	18	190	15.3	1008	3	US-08-987-121A-5	Sequence 5, Appli
	19	190	15.3	1011	3	US-08-987-121A-3	Sequence 3, Appli
	20	189	15.2	822	4	US-09-710-279-727	Sequence 727, App
	21	189	15.2	3993	4	US-09-710-279-3985	Sequence 3985, Ap
	22	187	15.1	1155	4	US-09-602-777A-147	Sequence 147, App
	23	186.5	15.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	24	186.5	15.0	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	25	186	15.0	1006	3	US-08-961-083-51	Sequence 51, Appli
	26	186	15.0	1006	3	US-09-536-784-51	Sequence 51, Appli
	27	186	15.0	1107	3	US-09-134-001C-1072	Sequence 1072, Ap
	28	186	15.0	10974	3	US-08-961-527-214	Sequence 214, App
	29	184	14.8	1011	3	US-09-066-512-1	Sequence 1, Appli
	30	182	14.7	1011	4	US-09-583-110-2196	Sequence 2196, Ap
	31	182	14.7	1663	4	US-09-620-312D-6	Sequence 6, Appli
	32	181.5	14.6	432	4	US-09-328-352-261	Sequence 261, App
	33	180.5	14.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
	34	180.5	14.6	1664976	4	US-09-692-570-1	Sequence 1, Appli
	35	178	14.4	640681	4	US-09-790-988-1	Sequence 1, Appli
	36	177	14.3	1011	4	US-09-107-433-1618	Sequence 1618, Ap
	37	174.5	14.1	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	38	174.5	14.1	1230230	4	US-09-438-185A-1	Sequence 1, Appli
	39	174	14.0	1026	3	US-09-149-624-1	Sequence 1, Appli
	40	173	14.0	1101	4	US-09-134-000C-1551	Sequence 1551, Ap
	41	170	13.7	15249	4	US-08-956-171E-102	Sequence 102, App
	42	170	13.7	15249	4	US-08-781-986A-102	Sequence 102, App
	43	168	13.5	3064	3	US-09-221-017B-794	Sequence 794, App
	44	159.5	12.9	42325	4	US-08-311-731A-111	Sequence 131, App
	45	145.5	11.7	3215	4	US-09-710-279-3566	Sequence 3566, Ap

## ALIGNMENTS

RESULT 1  
US-09-774-528-177  
; Sequence 177, Application US/09774528  
; Patent No. 6743619  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Zhou, Ping  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Zhang, Jie  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Xue, Aiding J.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6743619e1 Nucleic Acids and  
; FILE OF INVENTION: Polypeptides  
; FILE REFERENCE: 802  
; CURRENT APPLICATION NUMBER: US/09/774,528  
; NUMBER OF SEQ ID NOS: 441  
; SOFTWARE: pc\_fl\_genes Version 2.0  
; SEQ ID NO 177  
; LENGTH: 1416  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (205)..(1305)  
US-09-774-528-177

Alignment Scores:

Pred. No.: 1.02e-157 Length: 1416  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2\_copy\_176\_414 (1-239) x US-09-774-528-177 (1-1416)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB CTGTTGGCATTAAGTTCAGAGGTTTCAGATTTCCTGCTTGGAAAGCTTTGGACATA 645
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgLysSerLeuIleLysHisProGlu 40
DB GCACCAAGGTGACATCTTGACAGAGGTGGCAAGAGACTTTCTTAATAAATCAATCCAGAG 705
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB TGCTCCACATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 765
QY 61 HisPheAspIleLysProProLeuHisAlaLysAsnCysAspPheSerPheThrGly 80
DB CATTTTGACATCAAACTCCCTTGATCATGCTAAATTTGATTTTCTTTACTGGA 825
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 100
DB CTTCACACAGTATGATTAATAATATGAAAAAGAGAAAGAGATTTGGAAG 885
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB GGGCAAAATCCGTCTTCAGCAGCAGACATGCTGCACAGTACAGCAGCAATGGCAT 945
QY 886 GGGCAAAATCCGTCTTCAGCAGCAGACATGCTGCACAGTACAGCAGCAATGGCAT 945
DB CATCTTGGAAGAAAGACATGGGCTATTCTGTTTGTAGACAGAGACTTTTACT 1005
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 946 CATCTTGGAAGAAAGACATGGGCTATTCTGTTTGTAGACAGAGACTTTTACT 1005
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160
DB 1006 CAAAATAATGCAATGCTGTGATCTGTGTGTCGCAAGTAACTTCTAATATCCGCA 1065
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1066 GCTCTGGAATTTTAACAAGCAGACAGTGCATTTGTGTCTCTCCACAGACTA 1125
QY 181 CysThrAspAsnGlyIleMetIleAlaTPasnglyIleGluArgLeuArgAlaGlyLeu 200
DB 1126 TGCACTGATTAATGCGATTATGATTCATGCAATGATGAAAGACTACGTGCTG 1185
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrgluProLysCysProLeuGlyValAsp 220
DB 1186 GGCATTTTACATGATGAGATGAGAGCATCCGCTATGAAACCAAAATGCTCTTGGAGTAGAC 1245
QY 221 IleSerLysGluValGlyGlnLysSerIleLysValProGlnLeuLysMetGluIle 239
DB 1246 ATATCAAAAGAAAGTTGAGAGACTTCCATTAAGTACCAATTAATAAATGAGATA 1302

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
```

SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 23  
; LENGTH: 1526  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-067-443-23

Alignment Scores:

Pred. No.: 1.15e-157 Length: 1526  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-649-273-2\_copy\_176\_414 (1-239) x US-10-067-443-23 (1-1526)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB CTGTTGGCATTAAGTTCAGAGGTTTCAGATTTCCTGCTTGGAAAGCTTTGGACATA 144
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgLysSerLeuIleLysHisProGlu 40
DB 145 GCACCAAGGTGACATCTTGACAGAGGTGGCAAGAGACTTTCTTAATAAATCAATCCAGAG 204
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 205 TGCTCCACATGAGTGTGGGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 264
QY 61 HisPheAspIleLysProProLeuHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 265 CATTTTGACATCAAACTCCCTTGATCATGCTAAATTTGATTTTCTTTACTGGA 324
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 100
DB 325 CTTCACACAGTATGATTAATAATATGAAAAAGAGAAAGAGTATTTGAGAG 384
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB 385 GGGCAAAATCCGTCTTCAGCAGCAGACATGCTGCACAGTACAGCAGCAATGGCAT 444
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
DB 445 CATCTTGGAAGAAAGACATGGGCTATTCTTGTGTAGACAGAGACTTTGATCCT 504
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160
DB 505 CAAAATAATGCAATGCTGTGATCTGTGTGTCGCAAGTAACTTCTAATATCCGCA 564
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 565 GCTCTGGAATTTTAACAAGCAGACAGTGCATTTGTGTCTCTCCACAGACTA 624
QY 181 CysThrAspAsnGlyIleMetIleAlaTPasnglyIleGluArgLeuArgAlaGlyLeu 200
DB 625 TGCACTGATTAATGCGATTATGATTCATGCAATGATGAAAGACTACGTGCTG 684
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrgluProLysCysProLeuGlyValAsp 220
DB 685 GGCATTTTACATGATGAGAGCATCCGCTATGAAACCAAAATGCTCTTGGAGTAGAC 744
QY 221 IleSerLysGluValGlyGlnLysSerIleLysValProGlnLeuLysMetGluIle 239
DB 745 ATATCAAAAGAAAGTTGAGAGACTTCCATTAAGTACCAATTAATAAATGAGATA 801

RESULT 3
US-10-067-443-1
; Sequence 1, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: SPINAL CORD, MP-1
```

```
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (231)..(1472)
; US-10-067-443-1

Alignment Scores:
Pred. No.: 2,07e-157 Length: 2197
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-1 (1-2197)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAapPheLeuLeuGlyLysSerLeuAapIle 20
DB 756 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCGTTCTTGGAAAGCTTTGACATA 815
QY 21 AlaProGlyAaspMetLeuAapLysValAlaArgLeuSerLeuIleLysHisProGlu 40
DB 816 GCACCAAGTGCATGCTTGACAGAGTGCAGAAACCTTCTTAATAAACATCCAGAG 875
QY 41 CysSerThrmcSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnLysAapRph 60
DB 876 TGCTCCACCATGATGGTGGGAAAGCCATAGAACCTTTGGCCAAACAGAAATGATTT 935
QY 61 HisPheAapIleLysProProLeuHisHisAlaLysAapCysAapPheSerPheThrgly 80
DB 936 CATTTTGCATCAAACTCCCTTGATCATGCTAAATAATTTGATTTTCTTTTACTGGA 995
QY 81 LeuGlnHisValThrAapLysIleIleMetLysLysGlyLysGlnGlyIleGluLys 100
DB 996 CTTCAACACGTTACGTATTAATAATATATGAAAAAGAAAGAAAGATTTTGAAGA 1055
QY 101 GlnGlnIleLeuSerSerAlaAlaAapIleAlaIleThrValGlnHisThrmcAlaCys 120
DB 1056 GGGCAAAATCCGCTTTCAGACAGACATCTGTCACAGTACACACACAAATGGCATGT 1115
QY 121 HisLeuValLysAapThrmHisArgAlaIleLeuPheCysLysGlnArgAapLeuAapPro 140
DB 1116 CATCTTGTGAAGAAACACATCGGCTATTTCTGTTTGAAGACAGAGACTTGTTACTT 1175
QY 141 GlnAapAapAlaValLeuValAlaSerGlyValAlaAapAapPheThrglyIleArgArg 160
DB 1176 CAAATATATCAGATGCTGTGTCATCTGTGTGTGTCGCAAGTAACTTCAATATCCGAGA 1235
QY 161 AlaLeuGlnIleLeuThrAapAlaThrGlnCysThrmLeuLeuCysProProProArgLeu 180
DB 1236 GCTCTGAAATTTTAAACAAAGCAACAGATGCACTTTGTGTGCTCTCCCGAGACTA 1295
QY 181 CysThrmAapAapGlyIleMetIleAlaIleArgArgArgIleGluArgLeuArgAlaIleLys 200
DB 1296 TGCATCATATATGCACTTATATATGCAATGATATGAAAGCTACGAGCTGGCTTG 1355
QY 201 GlyIleLeuHisAapIleGlnGlyIleArgArgArgProLysCysProLeuGlnValAap 220
DB 1356 GGCATTTTACATGACATAGAGGATCGCTATGAAACCAAAATGCTCTTGAGATAGAC 1415
QY 221 IleSerLysGlnValGlyGlnAlaSerIleLysValProGlnLeuLysMetGlnIle 239
```

```
DB 1416 ATATCAAAAGAGTTGAGAAAGCTTCATTAAGTACCAAAATTAATAAGAGATA 1472
RESULT 4
US-10-067-443-21
; Sequence 21, Application US/10067443
; Patent No. 6642041
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
; US-10-067-443-21

Alignment Scores:
Pred. No.: 3.87e-154 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 0
Best Local Similarity: 90.15% Mismatches: 1
Query Match: 97.86% Indels: 25
DB: 4 Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-21 (1-1387)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAapPheLeuLeuGlnLysSerLeuAapIle 20
DB 549 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCGTTCTTGGAAAGCTTTGACATA 608
QY 21 AlaProGlyAaspMetLeuAapLysValAlaArgLeuSerLeuIleLysHisProGlu 40
DB 609 GCACCAAGTGCATGCTTGACAGAGTGCAGAAACCTTCTTAATAAACATCCAGAG 668
QY 41 CysSerThrmcSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnLysAapRph 60
DB 669 TGCTCCACCATGATGGTGGGAAAGCCATAGAACCTTTGGCCAAACAGAAATGATTT 728
QY 61 HisPheAapIleLysProProLeuHisHisAlaLysAapCysAapPheSerPheThrgly 80
DB 729 CATTTTGAATCAAACTCCCTTGATCATGCTAAATAATTTGATTTTCTTTTACTGGA 788
QY 81 LeuGlnHisValThrAapLysIleIleMetLysLysGlyLysGlnGlyIleGluLys 98
DB 788 CTTCAACACGTTACGTATTAATAATATGAAAAAGAAAGAAAGATATTTCTTA 848
QY 98 ----- 98
DB 849 ATTAGTAAGTTGAACAGATAATATATTCCTGATGTCCTAATAATAGCTGCTATTTC 908
QY 99 -----GlnLysGlyGlnIleLeuSerSerAlaAlaAapIleAlaIleThrValGln 115
DB 909 TGCAGGTATGAAAGGGCAAAATCCGCTTTCAGACAGACATCTGCTCCACAGTACG 968
QY 116 HisThrmcAlaCysHisLeuValLysAapThrmHisArgAlaIleLeuPheCysLysGln 135
DB 969 CACACAAATGGCATGCTTGTGAAGAAACATCGGCTATTTGTGTTGAAGCAG 1028
QY 136 ArgAapLeuLeuPheGlnAapAapValLeuValAlaSerGlyValAlaAapAap 155
DB 1029 AGAGACTTGTATACCTCAAAATATATGCAATGCTGTTGATGCTGTGCTCCAGTATAC 1088
QY 156 PheThrglyIleArgArgAlaLeuGlnIleLeuThrAapAlaThrGlnCysThrmLeuLys 175
```





Db	13220	TCGCGAAGTAATTCTGATATATCGGACAGGCTCTGGAAATTTTAAACAAGGACACAGG	13279
Qy	171	sthrleuleuCySProProProArgleuCySthAspansglyllewctlleaIaTyr--	190
Db	13280	CACCTTGTGTGTCTCTCTCCGACACTATGACATGATATATGATGATTCATGTGA	13339
Qy	190	-----	190
Db	13340	AGCCACAGATATACGTGCTTCACTACATPAACTATATTAATTGCCATTTATCAT	13399
Qy	190	-----	190
Db	13400	ACTAAGCCTTCCTCCCTTCACATCTTGAGCTATGATTTTATTAATGCTTCTATTTA	13455
Qy	191	--AsnrglyllegluaArgleuArgAlaGlyleuHIsAspIleuglyllea	210
Db	13460	GGAATGTATTTGAAAGACTACGTGCTGCTGGCATTTTACATGACATAGAGGCATCC	13519
Qy	210	rgTyrGluProLys	214
Db	13520	GCTATGAACCAAG	13533
RESULT 6			
US-09-596-002-38			
; Sequence 38, Application US/09596002			
; Patent No. 6632636			
; GENERAL INFORMATION:			
APPLICANT: Lagace, Robert, E.			
APPLICANT: Patterson, Chandra			
APPLICANT: Berg, Kim, L.			
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME			
FILE REFERENCE: PM-0008-4 US			
CURRENT APPLICATION NUMBER: US/09/596,002			
CURRENT FILING DATE: 2000-05-16			
PRIOR APPLICATION NUMBER: 60/140,121			
PRIOR FILING DATE: 1999-06-18			
NUMBER OF SEQ ID NOS: 41			
SOFTWARE: PERL Program			
SEQ ID NO 38			
LENGTH: 94750.			
TYPE: DNA			
ORGANISM: M. catarrhalis			
FEATURE:			
NAME/KEY: misc feature			
OTHER INFORMATION: incyte template ID No. 6632636 38			
PUBLICATION INFORMATION:			
US-09-596-002-38			

Alignment Scores:			
Pred. No.:	3, 02e-20	Length:	94750
Score:	248.50	Matches:	73
Percent Similarity:	48.67%	Conservative:	37
Best Local Similarity:	32.30%	Mismatches:	90
Query Match:	20.04%	Indels:	27
DB:	4	Gaps:	7
US-10-649-273-2_COPY_176_414 (1-239) x US-09-596-002-38 (1-94750)			
QY	1	LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyGlySerLeuAspIle	20
Db	24305	ATGCTGGTGGCCGCGCCATGCTGTGGCGGTATACGATATTGGCGACTCTATGATAT	24364
QY	21	AlaProGlyAspMetLeuAspIlyValAlaArgArgLeuSerLeuIleYshIProGlu	40
Db	24365	GCGGTGGGGATGCTTTGATTAACGGCAAAATGCTCAACG---CCCTATCT---	24418
QY	41	CysSerThrMetSerGlyGlyIlyValAlaIleGluHisLeuAlaIlyGlnGlyAspArgPhe	60
Db	24419	-----GGTGGCCCAAAATATGAAAAATATAGCCAAAAACGGCAACCCACAC	24463
QY	61	HisPheAspIleIlySerProPheLeuHisAlaIlyAsnIlyAspArgPheSerPheThrIly	80
Db	24464	GCGTATAGACTCGCAAGACCCACGACAT---AAAGCGCTGAGATTCTTCGTAGTGC	24520

Dy 1 LeuGlnHisValThrAspArgLysIleMetLeuLysGluLysGluGlyIleGluLys 100  
:::|||||  
24521 ATGAATACCGCCATTCAATATCTCATCAAGACACACCACAAAGCCCAAAGGACCCC--- 24577

Dd 1 GlyIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHstHmetAlaCys 120  
:::|||||  
24578 -----GCCAACGACGACATGCCTGGCAAGCTTTTGAGTATGCGGTGGTGGAT 24625

Dy 1 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnAraGlyLeuPro 140  
||| ||| |  
24626 ACTTGGTCATAAAAATGACACCAAGCATACAGATGACAGCATTCGCCAG----- 24676

Dd 1 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrlLeuArg 160  
||| ||| |  
24677 -----CTGGTGTGTCGACGGGGGGGTCTCTCCCAATCAGATCTACGGCGC 24721

Dy 1 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThreuleucysProProProArgLeu 180  
||| ||| |  
24722 ACCCTGACCGAGACGCTCGCCCAATCGATCGGTGTACTATGCCCGACCGAGCTTA 24781

Dy 1 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluLysLeuArgAlaGlyLeu 200  
||| ||| |  
24782 TGCAACGATTAATGTCGTGCGATGATCGCTATCGCTTTGTGGCGCTCACGTGGACAG 24841

Dd 1 GlyIleLeuHisAspIleGluGlyIleArgTyrlGluProLysCysProLeuGlyValAsp 220  
||| ||| |  
24842 TCG-----GATGACTGTGCG-GGTTGCTGTATTC-----CCGATGGGATATGAC 24885

Dy 1 IleSerLysGluValGly 226  
||| ||| |  
24886 GACGCTTGCGCGTATCGGC 24903

```

RESULT 7
US-09-540-236-806
; Sequence 806 Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 806
LENGTH: 1053
TYPE: DNA
ORGANISM: M.catarrhalis
US-09-540-236-806

Alignment Scores:
Pred. No.:          3.22e-23      Length:        1053
Score:              247.00       Matches:         70
Percent Similarity: 46.85%     Conservative:   34
Best local Similarity: 31.53%    Mismatches:     80
Query Match:        19.92%      Indels:         38
DB:                  Gaps:           6

US-10-649-273-2_COPY_176_414 (1-239) x US-09-540-236-806 (1-1053)
QY      1 LeuleuaIaleValaGInGIyVaISerAaspHeleuleuLeugLIyLSerLeuAspIle 20
Dbb:::|||||:::|||||:::|||||:::|||||
445 ATGTCTGGTCGCAGTCCGAATGTCTGTACAAATTATTTGGCGCAGTCATCATGATGAT 504
DY      21 ALaProGlyaaPmetLeuaspLysValaAlaaGArgLeuSerLeuileuLyshIsProGU 40
Dbb|||:||:|||||:::|||||:::|||||
505 GGCGTGCGCAATGCTTTGATTAACGCCAAAATGCTCAAATG---CCCTATCT-- 558
QY      41 CysSerThrMetSecGLyLYsaLaIEgLuHisLeuAlaysgInGIyaanAArgPhe 60
Db-----GGTGCTCCAATATCGAAAAATAAGCCAAAACGGGCAACCACAC 603

```

```
Qy 61 HispAspIleLeuProProLeuHisAlaIysAsnCyAspPheSerPheThrGly 80
Db 604 GCCATATAGCTGCCAAGACCGATGAGCAT---AAAGGCTGGATTTTTCGTTCAAGTGC 660
Qy 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100
Db 661 ATGAAACCGGCATTCATATCTCATCAAGACACGCCCAACGCCCAAGCAACCT--- 717
Qy 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleValGlnHisIleThrAlaCys 120
Db 718 -----GCCACAGACAGACATCGCCGAGCTTTGAGATCGGTTGGTGGAT 765
Qy 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
Db 766 ACTTTGGTCAAAAATGACACCAAGCACTACAGATGACAGCATTCGCCAG----- 816
Qy 141 GlnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheThrIleArgArg 160
Db 817 -----CTGGTGGTCCAGGGGGGCGTCTGCAATCAGACGCTACGCCCGC 861
Qy 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrIleLeuCysProProArgLeu 180
Db 862 ACCCTGACCGAGACCGCTCGCAAAATCGATGCGGTGGTGTACTACGCCCGACCGAGCTA 921
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGly--- 199
Db 922 TGCAACGATATATGTCGCATGATTCCTTACGCTGGCTTTTTCGCGCTAAGCCGTGGCAG 981
Qy 200 -----LeuGlyIleLeu 203
Db 982 TCGGATGACTTGGCGGTTGCTGTATTCGCCGATGGATATGACGATGCTGTATCGAA 1041
Qy 204 HisAsp 205
Db 1042 TATGAT 1047

RESULT 8
US-09-252-991A-884/C
; Sequence 884, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 884
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-884

Alignment Scores:
Pred. No.: 4,816-21 Length: 1059
Score: 231.00 Matches: 73
Percent Similarity: 49.54% Conservative: 34
Best Local Similarity: 33.80% Mismatches: 87
Query Match: 18.63% Indels: 22
DB: 4 Gaps: 8

US-10-649-273-2_COPY_176_414 (1-239) x US-09-252-991A-884 (1-1059)
Qy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
Db 621 TTGGTGGCGGTGACGGTATCGCGCTACAGTTGCTTGGCGAAATCGGTGACGATGCC 562
Qy 22 ProGlyAspMetLeuAspLysValAlaArgLeuSerLeuIleLysHisProGluCys 41
```

```
Db 561 GCCGGGAAGCCTTCGACAAAGCCGCAAGCTGATGGCCTG---GGTATATCC----- 511
Qy 42 SerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnIleLysAsnArgPheHis 61
Db 510 -----GGTGTCCGGAATAATGCCCGCTCGGAGAGCGGCACCTCTGGCGCC 463
Qy 62 PheAspIleLysProProLeuHisAlaIysAsnCyAspPheSerPheThrGlyLeu 81
Db 462 TTTCGTTCCTCCCGCGCGCATGACCGATCCGCCGCTGAGACTTTCAGCTTCACGGGCTC 403
Qy 82 GlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLysGly 101
Db 402 AAGACCTTTAACCTTGAAAC---ACCTGACAGCGTGGCTGAGAGCGCGGACGACAGCAG 346
Qy 102 GlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisIleThrAlaCysHis 121
Db 345 CAG-----ACCGCTGCGACATCGCCCTGCGCTTCAGACCCGCGTGTGAGACC 295
Qy 122 LeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGln 141
Db 294 CTGCTGATCAAGTCCGCTGCGCTTG-----AAGCAAGCCGCGCTG-----AAG 250
Qy 142 AsnAsnAlaValLeuValAlaSerGlyValAlaSerAspPheThrIleArgArgAla 161
Db 249 AAC-----CTGTGATTCGCGCGCGGTGTGACGCCAACACGAGCGCTGCCAGCGCGC 199
Qy 162 LeuGlnIleLeuThrAsnAlaThrGlnCysThrIleLeuCysProProArgLeuCys 181
Db 198 CTGGAAGAAGATGCTCGCGCAAAATGAAAGGGCAGGTGTTCTACGCCCGCGCTTCTGC 139
Qy 182 ThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeuGly 201
Db 138 ACCGACATGCGCGCATGATCGCTACGCCGCTGCCAGCGCTGCTCGCGGC----- 85
Qy 202 IleLeuHisAspIleGluGlyIleArgTyrgLysProLysCysProLeu 217
Db 84 ----CAGCATGACGCGCGCGCATCAGCGTCCAGCGCGCGCTGCCGATG 40

RESULT 9
US-09-252-991A-801
; Sequence 801, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 801
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-801

Alignment Scores:
Pred. No.: 5,956-21 Length: 1206
Score: 231.00 Matches: 73
Percent Similarity: 49.54% Conservative: 34
Best Local Similarity: 33.80% Mismatches: 87
Query Match: 18.63% Indels: 22
DB: 4 Gaps: 8

US-10-649-273-2_COPY_176_414 (1-239) x US-09-252-991A-801 (1-1206)
Qy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21
Db 621 TTGGTGGCGGTGACGGTATCGCGCTACAGTTGCTTGGCGAAATCGGTGACGATGCC 562
```

Db 604 TTGGTCGGGGTGAACGGATCGCGCGCTACCACTTGCTTGGCGAATCGATGACGATGCC 663  
Qy 22 ProglYAspMetLeuAspLysValAlaArgArgSerLeuIleLysHisProGluCys 41  
Db 664 GCCGGCGAAGCTTTCGACACAGCCGCAAGCTGATCGGCTG--GGCTATCC----- 714  
Qy 42 SerThrmSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHis 61  
Db 715 -----GGTGGTCGGGAATGCGCGCGCTCGGAGAGCGGCGACTCTCGCGGC 762  
Qy 62 PheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrcGlyLeu 81  
Db 763 TTGGTGTTCCTCCGCGCGATCGCATCGCCCGCGCTGAGCTTCAGCTTCAGCGGCGCTC 822  
Qy 82 GlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLysGly 101  
Db 823 AAGACCTTTCCTGAAAC--ACCTGGCAGCGCTTTCGTCGCGCGCGAGCGGACGACGCGAG 879  
Qy 102 GlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrmMetLacYsnHis 121  
Db 880 CAG-----ACCGCGTCGACATCGCCCTGGCGCTCCAGACCGCGGTGTCGAGAC 930  
Qy 122 LeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGln 141  
Db 931 CTGCTGATCAAGTCCGCTCGCGCTTG-----AAGCAGACCGGCTG-----AAG 975  
Qy 142 AsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrcIleArgArgAla 161  
Db 976 AAC-----CTGGTGAATCGCGCGGTGTCAGCGGCAACAGCGCTGCGCGCGGC 1026  
Qy 162 LeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProPheArgLeuCys 181  
Db 1027 CTGGAAAGATGCTCGCGCGAATGAGGCGCAGGTTCCTACGCCCGCGCTCTTCTGC 1086  
Qy 182 ThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeuGly 201  
Db 1087 ACCGACATGCGCGGATGATGATGCTAGCGCGGTCCAGCGCGCTGCTCGCGCGC----- 1140  
Qy 202 IleLeuHisAspIleGluGlyIleArgGlyLeuProLysCysProLeu 217  
Db 1141 ----CAGCATGACGCGCGCGATGACGCGTCCAGCGCGCTGCGCGCATG 1185  
RESULT 10  
US-09-557-884-1/c  
; Sequence 1, Application US/09557884  
; Patent No. 6506581  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 inch diskette  
; COMPUTER: Dell Pentium  
; OPERATING SYSTEM: MS DOS v6.22  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; FILING DATE: 25-Apr-2000  
; APPLICATION NUMBER: US/09/557,884  
; CLASSIFICATION: <Unknown>  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/476,102  
; FILING DATE: JUN-5-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB186P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504  
TELEFAX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-557-884-1  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
Alignment Scores:  
Pred. No.: 7.8e-15 Length: 1830121  
Score: 224.00 Matches: 67  
Percent Similarity: 47.50% Conservative: 28  
Best Local Similarity: 33.50% Mismatches: 83  
Query Match: 18.06% Indels: 22  
Gaps: 6  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-557-884-1 (1-1830121)  
Qy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21  
Db 552791 TTAGTCCGCTGATGATGCTAGGAAATATGAAAGTATGAGAAATCTATGATGATGCT 552732  
Qy 22 ProglYAspMetLeuAspLysValAlaArgArgSerLeuIleLysHisProGluCys 41  
Db 552731 GGTGGGAAGCCTTGATTAACAGCAAAATTAATCTGACTA---GATTATCA----- 552681  
Qy 42 SerThrmSerGlyGlyValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHis 61  
Db 552680 -----GGTGGCGCGCGACCTTCTCGTTAGCGGAAAGATGCGCAAAATCGT 552633  
Qy 62 PheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrcGlyLeu 81  
Db 552632 TTCAACATTTCCAGCTTCATGACAGATCGTCGACGCGCTTGAATTTTGTGCTTTA 552573  
Qy 82 GlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGly-----IleGlu 99  
Db 552572 AAAACATTTGCGCGCAATACATTAATCAAGCAATTAAGAGGCGGCACTGATGAG 552513  
Qy 100 LysGlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrmMetAla 119  
Db 552512 CAA-----ACTAAACAGATATTGCTTATGCTTCCAGATGCGGCGTG 552468  
Qy 120 CysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeu 139  
Db 552467 GATACCTTGGC-----ATTAAATGTAAAG---CGTGCATTGAAA 552432  
Qy 140 ProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrcIleArg 159  
Db 552431 GAAACAGGCTATTAACGTTATGATGCGGAGGCGGTAGCGCAATTAAGAACTCCCA 552372  
Qy 160 ArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysAspProProArg 179  
Db 552371 GAAACGCTTGGCACATTATGCAAAATTAGGTGGGAGGATTTTATCTCAACTCA 552312  
Qy 180 LeuCysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGly 199  
Db 552311 TTTTGTACAGATTAATGTCGATGATGCTTACACAGGTTTTTTACGTTTAAACAAAGT 552252  
RESULT 11  
US-09-643-990A-1/c  
; Sequence 1, Application US/09643990A  
; Patent No. 6528289  
; GENERAL INFORMATION:  
; APPLICANT: Robert D. Fleischmann  
; Mark D. Adams  
; Owen White  
; Hamilton O. Smith

J. Craig Venter  
TITLE OF INVENTION: The Nucleotide sequence of  
the Haemophilus influenzae Rd Genome, Fragments  
thereof, and Uses Thereof  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville,  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A  
FILING DATE: 23-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB186P1C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1  
Alignment Scores:  
Pred. No.: 7.8e-15 Length: 1830121  
Score: 224.00 Matches: 67  
Percent Similarity: 47.50% Conservative: 28  
Best Local Similarity: 33.50% Mismatches: 83  
Query Match: 18.06% Indels: 22  
Gaps: 4  
DB: 6  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-643-990A-1 (1-1830121)  
QY 2 LeuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21  
DB 552791 TTACTGGGTGTCAGTGGTGTGGAATAATAGAGATAGAGATATGATATGAT 552732  
QY 22 ProGlyAspMetLeuAspIleValAlaArgLysSerLeuIleLysHisProGlyCys 41  
DB 552731 GCTGGCGAGCCTTGTGATTAACAGCAAAATTAATCTTGACCTA-----GATTATCA----- 552681  
QY 42 SerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGInGlyAsnArgPheHis 61  
DB 552680 -----GGTGGCGCGGCACTTCTCGTTTAGCGGAAAAAGTACGCCCAATCGT 552633  
QY 62 PheAspIleLysProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGlyLeu 81  
DB 552632 TTTCACATTTCCACGTCGCAATACAGATGTCGAGCCCTTGATTTAGTTTCTCGTTTA 552573  
QY 82 GlnHisValThrAspLysIleIleMetLysGlyGlyGlyGly-----IleGlu 99  
DB 552572 AAACATTTGCGCGCAATACAGTTTAATCAAGCAATTAATAAGAGGCGAAGCTGATAGAG 552513  
QY 100 LysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAla 119

DB 552512 CAA-----ACTAACAGATATGCTTATGCTTCCAAAGATGCGGTGG 552468  
QY 120 CysHisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeu 139  
DB 552467 GATACTCTTGGC-----ATTAATGTAG--CGTGCATTGAAA 552432  
QY 140 ProGlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArg 159  
DB 552431 GAAACAGGCTATTAACCGTTTATGATTCGGAGAGGGGTAGCGCAATATAAACTCCCA 552372  
QY 160 ArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysAspProPheArg 179  
DB 552371 GAAACGCTTGCGCACTTATATGCAAAATTAGGTGGGAGAGTGTATTATCTCAACTCA 552312  
QY 180 LeuCysThrAspAsnGlyIleMetIleAlaThrPheGlnGlyIleGluArgLeuArgAlaGly 199  
DB 552311 TTTTGTACAGATTAATGTCGATGATGCTTAACAGGGTTTATTAAGTTTAAACAAAGGT 552252  
RESULT 12  
US-09-902-540-6612  
Sequence 6612, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 6612  
LENGTH: 996  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-6612  
Alignment Scores:  
Pred. No.: 8.47e-20 Length: 996  
Score: 221.50 Matches: 70  
Percent Similarity: 47.50% Conservative: 25  
Best Local Similarity: 35.00% Mismatches: 74  
Query Match: 17.86% Indels: 31  
Gaps: 4  
DB: 8  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-902-540-6612 (1-996)  
QY 5 ValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAlaProGlyAsp 24  
DB 433 GTGCAGGCTTACCGGAGTACCGGCTGTGGGACAGACGCGACGCGCGCGCGAG 492  
QY 25 MetLeuAspIleValAlaArgArgLeuSerLeuIleLysHisProGlyCysSerThrMet 44  
DB 493 GCATATGACAAACGCGCTGATCTCGGCGTGG--CCGATATCG----- 534  
QY 45 SerGlyGlyLysAlaIleGluHisLeuAlaLysGInGlyAsn-----ArgPhe 60  
DB 535 ---GGTGGCGAGCCCATGACACAGTTGGCGGACAGGGGAAACCGGAGCCATCCGCTTC 591  
QY 61 HisPheAspIleLysProProLeuHisIleAlaLysAsnCysAspPheSerPheThrGly 80  
DB 592 -----CCGGCGCGCTGCGGCGACAACTTGACGTCGTCCTCCGGG 636  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyGlyGlyGlyIleGluLys 100  
DB 637 TTGAAG-----ACGGCGGTGCTGCACCAAGTGCAGAAAGCAGCGCGCGAG 684  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120



Db 790 TTGGCAATTAATATGTCGACGA-----TTAGAGCAA 822  
Qy 142 AsnAen---AlaValLeuValAlaSerGlyValAlaSerAspPheTyrIleArg 160  
Db 823 ACAGGCTTTAAACGGTTAGATGCTGGGGCGTAAAGCTTAACCTTACCGCC 882  
Qy 161 AlaleuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
Db 883 AAAATGGCGATGATATGGAACAACCTCGAGGGGAAGTCTTTATGCTGCGCTGAGTTA 942  
Qy 181 CysThrAspAsnGlyIleMetIleAlaTPrAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db 943 TGTACCGATTAATGGCGCATGATGCTTGGCGGGATGATCGCTTTTAAAGGTGATACC 1002  
Qy 201 ---GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeu 217  
Db 1003 GAGGGGCGCTTTA-----GGGGTGACATGAGACACGTTGGCTTTA 1044

## RESULT 15

US-09-489-039A-2050  
; Sequence 2050, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2050  
; LENGTH: 1032  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2050

Alignment Scores:  
Pred. No.: 1.55e-17 Length: 1032  
Score: 205.00 Matches: 66  
Percent Similarity: 43.58% Conservative: 29  
Best Local Similarity: 30.28% Mismatches: 93  
Query Match: 16.53% Indels: 30  
DB: 4 Gaps: 7

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-09-489-039A-2050 (1-1032)

Qy 2 LeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIleAla 21  
Db 442 TTGATTACCGGTACCGGATATGGTCAATGAACTGCGGCGAATGACATTAACAATGCG 501  
Qy 22 ProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGluCys 41  
Db 502 GCGGGCGAAGCTTTGATTAAGACGCGGAAGCTGCGACTG---GATTATCC----- 552  
Qy 42 SerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPheHis 61  
Db 553 -----GGCGGGCGGCTGCTGCAAAATGCGCTCCAGGGCACCGAAGCGCGC 600  
Qy 62 PheAspIleLysProProLeuHisHisAlaLysAsnGlyAspPheSerPheThrGlyLeu 81  
Db 601 TTGTGCTTCCCGCGCGATACCGACCGTCCGGGGCTGGAAGCTTCAAGCTTTCGGGCTG 660  
Qy 82 GlnHisValThrAspLysIleIleMetLysGluLysGluGluGlyIleGluLysGly 101  
Db 661 AAGACCTTGGCGCGCAACACCATTCGACGCAACGCGGACGATGAG----- 705  
Qy 102 GlnIleuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCysHis 121  
Db 706 -----CAAAACCGCGCGACATCGCGCGGCGGCTTGAAGATGCGGTGCTGATACG 756  
Qy 122 LeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuProGln 141

Db 757 CTGATGATTAATATGTCGCGCGCG-----CTGAGACAA 789  
Qy 142 AsnAen---AlaValLeuValAlaSerGlyValAlaSerAspPheTyrIleArg 160  
Db 790 ACCGCTTTAAGCGTGTGTGATGCGGGAGGAGGCAACCGTAACCTGCGCGCG 849  
Qy 161 AlaleuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
Db 850 AAGCTGCGGAGATGATGCAAAACGCGCGGAGGTGTTTACGCGCGCTGACTTC 909  
Qy 181 CysThrAspAsnGlyIleMetIleAlaTPrAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db 910 TGTACTGACACAGCGCGCATGATCGCTACGCGCGATGATGCTGCTGCAACCGCGCGC 969  
Qy 201 GlyIleLeuHisAspIleGlu---GlyIleArgTyrGluProLysCysProLeu 217  
Db 970 -----AAAGCCGAGCTCGGCGTACGCTTGGCGCGCGCTGCGCGCTG 1011

Search completed: February 16, 2005, 22:38:34  
Job time : 1516.1 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 18:03:13 ; Search time 478 Seconds

(without alignments)  
2950.230 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414

Perfect score: 1240  
Sequence: 1 L1ALVQGVSPFLIGKSLDI.....DISKEVGSAIKVPOLKMEI 239

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 5378673 seqs, 2950229984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.epool/US10649273/runat\_14022005\_114706\_16497/app\_query.fasta\_1.1429  
-DB=Published Applications NA -OPMT=fastap -SUFFIX=nmph -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62  
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10649273@cgn\_1\_1\_1053@runat\_14022005\_114706\_16497  
-NCPU=6 -ICPU=3 -NO WMAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PTCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PTCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	1240	100.0	1416	17	US-10-120-988-177
2	1240	100.0	1526	14	US-10-067-443-23
3	1240	100.0	1526	17	US-10-649-273-23
4	1240	100.0	1526	17	US-10-651-722-23
5	1240	100.0	2197	14	US-10-067-443-1
6	1240	100.0	2197	17	US-10-649-273-1
7	1240	100.0	2197	17	US-10-651-722-1
8	1213.5	97.9	1387	14	US-10-067-443-21
9	1213.5	97.9	1387	17	US-10-649-273-21
10	1213.5	97.9	1387	17	US-10-651-722-21
11	1203	97.0	1245	14	US-10-012-140-6
12	1203	97.0	1820	14	US-10-012-140-4
13	1059	85.4	2208	17	US-10-094-749-400
14	1059	85.4	2890	18	US-10-723-860-7447
15	725	58.5	14364	14	US-10-067-443-20
16	725	58.5	14364	17	US-10-649-273-20
17	725	58.5	14364	17	US-10-651-722-20
18	468	37.7	371	18	US-10-430-201-3118
19	468	37.7	371	18	US-10-430-201-3119
20	337	27.2	1917	17	US-10-424-559-66417
21	270	21.8	1628	18	US-10-437-963-11249
22	257	20.7	1146	17	US-10-282-122A-11043
23	248.5	20.0	1000	18	US-10-343-561-50
24	248.5	20.0	1044	17	US-10-282-122A-26972
25	248.5	20.0	94750	17	US-10-672-787-38
26	247	19.9	936	17	US-10-282-122A-8315
27	241	19.4	1032	17	US-10-282-122A-11043
28	239	19.3	756	14	US-10-081-051-8
29	239	19.3	4360	14	US-10-081-051-2
30	231	18.6	1026	9	US-09-815-242-7701
31	231	18.6	1026	17	US-10-282-122A-30016
32	224	18.1	1029	9	US-09-815-242-6946
33	224	18.1	1029	17	US-10-282-122A-22020
34	224	18.1	1830121	17	US-10-329-670-1
35	224	18.1	1830121	18	US-10-158-865-1
36	213	17.2	1014	9	US-09-815-242-9682
37	212	17.2	1014	17	US-10-282-122A-39301
38	212	17.1	1020	17	US-10-282-122A-12254
39	210	16.9	1014	17	US-10-282-122A-11977
40	208	16.8	1007	17	US-10-282-122A-19220
41	208	16.8	1728	18	US-10-437-963-9137
42	207.5	16.7	1023	9	US-10-282-122A-31809
43	206	16.6	1014	17	US-09-815-242-6207
44	206	16.6	1014	17	US-10-282-122A-20511
45	206	16.6	1775	17	US-10-425-114-28614

#### ALIGNMENTS

RESULT 1  
US-10-120-988-177  
; Sequence 177, Application US/10120988  
; Publication No. US20030219745A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Ren, Feiyun  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 8020CN  
; CURRENT APPLICATION NUMBER: US/10/120,988  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 09/774,528  
; PRIOR FILING DATE: 2001-01-30  
; NUMBER OF SEQ ID NOS: 441  
; SOFTWARE: pt\_fl\_genes Version 2.0



```
; SEQ ID NO 177
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (205)..(1305)
US-10-120-988-177

Alignment Scores:
Pred. No.: 9,42e-152 Length: 1416
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-120-988-177 (1-1416)

QY 1 LeuLeuAlaLeuValAlaGlnGlyValSerAspPheLeuLeuGlyLeuSerLeuAspIle 20
DB CTGTTGGCATTAGTTCAAGAGAGTTTCAAGATTTCTGCTTTTGGAAGCTTTGACATTA 645
QY 21 AlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleYshHisProGlu 40
DB GCACCAAGTGACATGCTTGACAGGTGGCAAGAACCTTTCTTTATATAAATCCAGAG 705
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaYleGlnGlyAsnArgPhe 60
DB TGCTCCACCATGAGTGGTGGGAAACCATAGAACATTTGGCCAAACAGAAATAGATT 765
QY 61 HisPheAspIleYshProProLeuHisHisAlaYleAsnCysAspPheSerPheThrGly 80
DB CATTTTGACATCAAACTCCCTTGATCATGCTAAATAATGATTTTCTTTTACTGGA 825
QY 81 LeuGlnHisValThrAspIleYshIleMetIleYshGluYshGluGluGlyIleGluYsh 100
DB CTTCAACACGTTACTGATTAATTAATGAAAAAGAAAAAGAGAGTTTGAAGAG 885
QY 826 CTTCAACACGTTACTGATTAATTAATGAAAAAGAAAAAGAGAGTTTGAAGAG 885
DB 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB GGGCAAAATCTGTCTTGACAGACAGACATGCTGCCACAGTACACACATGGCATGT 945
QY 121 HisIleValIleYshArgThrHisArgAlaIleLeuPheCysYshGlnArgAspLeuPro 140
DB CATCTTGGAAGAAACACATCGGCTATCTGTTTGTGAAGACAGAGACTGTACTCT 1005
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheYrIleArgArg 160
DB CAAATAATATGACATGCTGTTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1065
QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB GCTCTGGAATTTTAAACAAACGACACACAGTGCCTTTGTTGTCTCTCCAGACTA 1125
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB TGCACTGATTAATGAGATTAATGATGATGAGAAATGATTAAGAACTACGCTGCTG 1185
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProYshCysProLeuGlyValAsp 220
DB GGCATTTTACATGACATTAAGAGCATCCGCTATGAAACCAAAATGTCTCTTGAGTAA 1245
QY 221 IleSerIySGluValGlyGluAlaSerIleYshValProGlnLeuYshMetGluIle 239
DB ATATCAAAAGAGGTGGAGAGCTTCCATTAAGATCAACATTAATAATGAGATTA 1302

RESULT 2
US-10-067-443-23
; Sequence 23, Application US/10067443
; Publication No. US20030082782A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
```

```
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEIN HIGHLY EXPRESSED IN
; FILE OF INVENTION: SPINUL CORP, MP-1
; FILE REFERENCE: D0073 NP
; CURRENT APPLICATION NUMBER: US/10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-067-443-23

Alignment Scores:
Pred. No.: 1.05e-151 Length: 1526
Score: 1240.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-067-443-23 (1-1526)

QY 1 LeuLeuAlaLeuValAlaGlnGlyValSerAspPheLeuLeuGlyLeuSerLeuAspIle 20
DB CTGTTGGCATTAGTTCAAGAGAGTTTCAAGATTTCTGCTTTTGGAAGCTTTGACATTA 144
QY 21 AlaProGlyAspMetLeuAspLeuValAlaArgArgLeuSerLeuIleYshHisProGlu 40
DB GCACCAAGTGACATGCTTGACAGGTGGCAAGAACCTTTCTTTATATAAATCCAGAG 204
QY 41 CysSerThrMetSerGlyGlyValAlaIleGluHisLeuAlaYleGlnGlyAsnArgPhe 60
DB TGCTCCACCATGAGTGGTGGGAAACCATAGAACATTTGGCCAAACAGAAATAGATT 264
QY 205 TGCTCCACCATGAGTGGTGGGAAACCATAGAACATTTGGCCAAACAGAAATAGATT 264
DB 61 HisPheAspIleYshProProLeuHisHisAlaYleAsnCysAspPheSerPheThrGly 80
DB CATTTTGACATCAAACTCCCTTGATCATGCTAAATAATGATTTTCTTTTACTGGA 324
QY 81 LeuGlnHisValThrAspIleYshIleMetIleYshGluYshGluGluGlyIleGluYsh 100
DB CTTCAACACGTTACTGATTAATTAATGAAAAAGAAAAAGAGAGTTTGAAGAG 384
QY 325 CTTCAACACGTTACTGATTAATTAATGAAAAAGAAAAAGAGAGTTTGAAGAG 384
DB 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
DB GGGCAAAATCTGTCTTGACAGACAGACATGCTGCCACAGTACACACATGGCATGT 444
QY 385 GGGCAAAATCTGTCTTGACAGACAGACATGCTGCCACAGTACACACATGGCATGT 444
DB 121 HisIleValIleYshArgThrHisArgAlaIleLeuPheCysYshGlnArgAspLeuPro 140
DB CATCTTGGAAGAAACACATCGGCTATCTGTTTGTGAAGACAGAGACTGTACTCT 504
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheYrIleArgArg 160
DB CAAATAATATGACATGCTGTTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 564
QY 505 CAAATAATATGACATGCTGTTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 564
DB 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB GCTCTGGAATTTTAAACAAACGACACACAGTGCCTTTGTTGTCTCTCCAGACTA 624
QY 181 CysThrAspAsnGlyIleMetIleAlaIleArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB TGCACTGATTAATGAGATTAATGATGATGAGAAATGATTAAGAACTACGCTGCTG 684
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProYshCysProLeuGlyValAsp 220
DB GGCATTTTACATGACATTAAGAGCATCCGCTATGAAACCAAAATGTCTCTTGAGTAA 744
QY 221 IleSerIySGluValGlyGluAlaSerIleYshValProGlnLeuYshMetGluIle 239
DB ATATCAAAAGAGGTGGAGAGCTTCCATTAAGATCAACATTAATAATGAGATTA
```

Db 745 ATATCAAAAGAGTTGAGAGCTTCATTAATAAGTACCAATTAATAATGGAGATA 801

RESULT 3

US-10-649-273-23

/ Sequence 23, Application US/10649273

/ Publication No. US20040043407A1

/ GENERAL INFORMATION:

/ APPLICANT: Bristol-Myers Squibb Company

/ TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1

/ FILE REFERENCE: D0073 CNT

/ CURRENT APPLICATION NUMBER: US/10/649,273

/ CURRENT FILING DATE: 2003-08-27

/ PRIOR APPLICATION NUMBER: US 60/266,518

/ PRIOR FILING DATE: 2001-02-05

/ PRIOR APPLICATION NUMBER: US 10/067,443

/ PRIOR FILING DATE: 2002-02-05

/ PRIOR APPLICATION NUMBER: US 60/282,814

/ PRIOR FILING DATE: 2001-04-10

/ NUMBER OF SEQ ID NOS: 71

/ SOFTWARE: PatentIn version 3.2

/ SEQ ID NO 23

/ LENGTH: 1526

/ TYPE: DNA

/ ORGANISM: homo sapiens

US-10-649-273-23

Alignment Scores:

Pred. No.:	1.05e-151	Length:	1526
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	17	Gaps:	0

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-649-273-23 (1-1526)

QY 1 LeuEua1aLeuValGInG1yValSerAppPheLeuLeuG1ySerLeuApp1le 20

Db 85 CTGTGGCATTTAGTTCAAGAGTTTCAGATTTTCGCTTTGGAAAAGCTTTGGACATA 144

QY 21 AlAProG1yAPMeLeuAPlySVa1A1aRgA1gLeuSerLeu11e1ySH1SProG1u 40

Db 145 GCACCAAGTGAATGCTTGAACAGGTGGCAAGAGACTTTCTTTAATAAACATCCAG 204

QY 41 CysSerThrMetSerG1yG1yVala1a1eG1uH1sLeuA1a1ySG1nG1yASnaRgPhe 60

Db 205 TGTCTCCACATGAGTGGGAAAAGCCATAGAACATTTGGCCAAAGAAATGATTT 264

QY 61 HisPheAP11e1ySProPheuH1SH1a1a1ySAsnCyASpPheSerPheThrg1y 80

Db 265 CATTTTGACATCAAACTCCCTTCATCATGCTAATAAATTTGATTTTCTTTTACTGA 324

QY 81 LeuG1nH1sVal1ThrAPlyS1e11eMe1yS1ySG1u1ySG1uG1y11eG1u1yS 100

Db 325 CTTCACACGCTTACGATTAATAATGAAGAAAAGAAAAGAGAGTATTGAGAG 384

QY 101 G1yG1n11e1ySerSer1a1a1aSP11e1a1a1ThrVal1G1nH1sThrMetA1CyS 120

Db 385 GGGCAAAATCCGTCTTCAGCAGACATGCTGCAAGTACAGACACAAATGGCATGT 444

QY 121 HisLeuVal1ySArgThrH1sArgA1a11eLeuPheCyS1ySG1nArgASpLeuPro 140

Db 445 CATCTTGTAAGAAAACACATCGGCTAATTCGTTTGTAAAGCAAGAGACTTTTACT 504

QY 141 G1ASnaSna1a1a1e1yVala1a1eSerG1yG1yVala1a1eSerASpPheThrg1y 160

Db 505 CAATAATATGACGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 564

QY 161 AlaleuG1u11e1yThrASna1a1Thrg1nCyS1ThrLeu1yCySProPheProArgLeu 180

Db 565 GCTCTGGAATTTTAAACAAAGCAACAGTGAATTTGTTGTGTCTCTCCAGACTA 624

QY 181 CysThrASpAng1y11eMe11e1a1a1ThrAng1y11eG1u1ySG1nArgA1a1a1yLeu 200

Db 625 TGCACGTATTAATGCAATTAATGATTCATGAAATGATTAAGAAAGTACGCTGCTTG 684

QY 201 G1y11e1yH1sAP11e1yG1u1y11eArg1yRg1uPro1yCySProLeuG1yVala1aP 220

Db 685 GGCATTTTATGACATGACATTAAGAGGCAATCCGCTATGAACCAAAATGCTCTTGGAGTAGAC 744

QY 221 11eSer1ySG1uValG1yG1u1a1eSer11e1ySVa1P1roG1nLeu1yMe1G1u1e 239

Db 745 ATATCAAAAGAGTTGAGAGACTTCATTAATAAGTACCAATTAATAATGGAGATA 801

RESULT 4

US-10-651-722-23

/ Sequence 23, Application US/10651722

/ Publication No. US20040048302A1

/ GENERAL INFORMATION:

/ APPLICANT: Bristol-Myers Squibb Company

/ TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALOPROTEASE, MP-1

/ FILE REFERENCE: D0073 DIV

/ CURRENT APPLICATION NUMBER: US/10/651,722

/ CURRENT FILING DATE: 2003-08-29

/ PRIOR APPLICATION NUMBER: US 60/266,518

/ PRIOR FILING DATE: 2001-02-05

/ PRIOR APPLICATION NUMBER: US 10/067,443

/ PRIOR FILING DATE: 2002-02-05

/ PRIOR APPLICATION NUMBER: US 60/282,814

/ NUMBER OF SEQ ID NOS: 71

/ SOFTWARE: PatentIn version 3.2

/ SEQ ID NO 23

/ LENGTH: 1526

/ TYPE: DNA

/ ORGANISM: homo sapiens

US-10-651-722-23

Alignment Scores:

Pred. No.:	1.05e-151	Length:	1526
Score:	1240.00	Matches:	239
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	17	Gaps:	0

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-651-722-23 (1-1526)

QY 1 LeuEua1aLeuValGInG1yValSerAppPheLeuLeuG1y1ySerLeuApp1le 20

Db 85 CTGTGGCATTTAGTTCAAGAGTTTCAGATTTTCGCTTTGGAAAAGCTTTGGACATA 144

QY 21 AlAProG1yAPMeLeuAPlySVa1A1aRgA1gLeuSerLeu11e1ySH1SProG1u 40

Db 145 GCACCAAGTGAATGCTTGAACAGGTGGCAAGAGACTTTCTTTAATAAACATCCAG 204

QY 41 CysSerThrMetSerG1yG1yVala1a1eG1uH1sLeuA1a1ySG1nG1yASnaRgPhe 60

Db 205 TGTCTCCACATGAGTGGGAAAAGCCATAGAACATTTGGCCAAAGAAATGATTT 264

QY 61 HisPheAP11e1ySProPheuH1SH1a1a1ySAsnCyASpPheSerPheThrg1y 80

Db 265 CATTTTGACATCAAACTCCCTTCATCATGCTAATAAATTTGATTTTCTTTTACTGA 324

QY 81 LeuG1nH1sVal1ThrAPlyS1e11eMe1yS1ySG1u1ySG1uG1y11eG1u1yS 100

Db 325 CTTCACACGCTTACGATTAATAATGAAGAAAAGAAAAGAGAGTATTGAGAG 384

QY 101 G1yG1n11e1ySerSer1a1a1aSP11e1a1a1ThrVal1G1nH1sThrMetA1CyS 120

Db 385 GGGCAAAATCCGTCTTCAGCAGACATGCTGCAAGTACAGACACAAATGGCATGT 444

QY 121 HisLeuVal1ySArgThrH1sArgA1a11eLeuPheCyS1ySG1nArgASpLeuPro 140

Db 445 CATCTTGTAAGAAAACACATCGGCTAATTCGTTTGTAAAGCAAGAGACTTTTACT 504

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTyrIleArgArg 160  
DB 505 CAATAATATGAGTACTGCTGTCATCTGCTGTCGCGAAGTAACTTCTATATCCGACA 564  
QY 161 AlaleuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlu 180  
DB 565 GCTCTGGAATTTTAAACAAACGACACAGTGCATCTTGTGTGTCCTCCACAGACTA 624  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 625 TGCCTGATTAATGCGATTATATGATGATGAAATGGTATTAAGAACTACGTCTGCTTG 684  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220  
DB 685 GGCAATTTACATGACATGAAAGGCACTCCGCTATGAACCAAAATGTCCTTGGAGTAAAC 744  
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
DB 745 ATATCAAAAGAAGTTGAGAAAGCTTCCATTAAGTACCAATTAATAATGAGATA 801

## RESULT 5

US-10-067-443-1  
; Sequence 1, Application US/10067443  
; Publication No. US20030082782A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED  
; FILE REFERENCE: D0073 NP  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/10/067,443  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 2197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (231)..(1472)  
US-10-067-443-1

## Alignment Scores:

Pred. No.: 1,79e-151 Length: 2197  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-067-443-1 (1-2197)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 756 CTGTTGGCATTTAGTTCAGAGATTTCAGATTTTCGCTTCTTGAAGAGCTTTGACATA 815  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 816 GCACCAAGGTGACATGCTTGACAGGTGCAAGAAACATTTCTTAAATAAACATCCAGG 875  
QY 41 CysSerThrMetSerGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 876 TGCCTCCACCAAGAGTGGTGGAAAGCCATAGACATTTGGCCAAACAGAGAAATGATT 935  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
DB 936 CATTTTGCATCAAACTCCCTTGATCATGCTTAATAATTTGATTTTCTTTACTGGA 995  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGluGlyIleGluLys 100

DB 996 CTTCAACACGTTACTGATTAATTAATTAAGAAAAAGAGAGATATTGAGAG 1055  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAlaThrValGlnHisThrMetAlaCys 120  
DB 1056 GGGCAAAATCTCTCTTACAGACAGACAGATGCTGCGCACAGTACAGACACAAATGGCATGT 1115  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGluArgAspLeuPro 140  
DB 1116 CATCTTGAAAAAGAACATCGGCTATTCGTTTGTGAAGAGAGAGATTTTACCT 1175  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAspPheTyrIleArgArg 160  
DB 1176 CAATAATATGAGTACTGCTGTCATCTGCTGTCGCGAAGTAACTTCTATATCCGACA 1235  
QY 161 AlaleuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProAlaGlu 180  
DB 1236 GCTCTGGAATTTTAAACAAACGACACAGTGCATCTTGTGTGTCCTCCACAGACTA 1295  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 1296 TGCCTGATTAATGCGATTATATGATGATGAAATGGTATTAAGAACTACGTCTGCTTG 1355  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220  
DB 1356 GGCAATTTACATGACATGAAAGGCACTCCGCTATGAACCAAAATGTCCTTGGAGTAAAC 1415  
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
DB 1416 ATATCAAAAGAAGTTGAGAAAGCTTCCATTAAGTACCAATTAATAATGAGATA 1472

## RESULT 6

US-10-649-273-1  
; Sequence 1, Application US/10649273  
; Publication No. US20040043407A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1  
; FILE REFERENCE: D0073 CNT  
; CURRENT FILING DATE: US/10/649,273  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2003-08-27  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 10/067,443  
; PRIOR FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/282,814  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 2197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (231)..(1472)  
US-10-649-273-1

## Alignment Scores:

Pred. No.: 1,79e-151 Length: 2197  
Score: 1240.00 Matches: 239  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 17 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-649-273-1 (1-2197)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 756 CTGTTGGCATTTAGTTCAGAGATTTCAGATTTTCGCTTCTTGAAGAGCTTTGACATA 815  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40

Db	816	GCACGAGTGACATCTTGACAGAGGTGGCAAGAGACTTTCTTTATATAAACATCCAGG	875
Qy	41	CysSerThrMetSerGlyValLeuAlaIleGluHisLeuAlaIleGlnGlyAsnArgPhe	60
Db	876	TGCTCCACCAAGAGGTGGGGAACCCATTAGAACATTTGGCCAAACAGAGAAATGATTT	935
Qy	61	HisPheAspIleValProProLeuHisHisAlaIleAsnGlySerPheSerPheThrGly	80
Db	936	CATTTTGACATCAACCTCCCTGATCATGCTAAATAATTTGATTTTCTTTTAACTGGA	995
Qy	81	LeuGlnHisValThrAspLeuIleIleMetLysLysGlyValGluGlyIleGluLys	100
Db	996	CTTCACACACGCTTACGTATTAATTAATTAATGAAAAAGAAAAAGAGAAAGTATGGAGAG	1055
Qy	101	GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys	120
Db	1056	GGGCAAAATCCGTGCTTTCAGACAGACAGACTTGCTGCCACAGTACAGACACAATGGCATGT	1115
Qy	121	HisLeuValIleValArgThrHisArgAlaIleLeuPheCysLysGlnAlaGlyAspLeuPro	140
Db	1116	CATCTTGGAAGAAAGAACACATCGGGCTATTCGTGTTTGTAAAGACAGAGACTTGTACTT	1175
Qy	141	GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg	160
Db	1176	CAAAATATATGACAGTACTGTGTCATCTGTGTGTGTGCGAATTAATCTTATATCCGACGA	1235
Qy	161	AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu	180
Db	1236	GCTCGAATAATTTTAAACAAGCAACACAGTGCACCTTTGTTGTGTGCTCCACGACTA	1295
Qy	181	CysThrAspAsnGlyIleMetIleAlaIleArgGlyIleGluValGluLeuArgAlaGlyLeu	200
Db	1296	TGCACGTGTAATGCGATTATGATTCATGAGATGGTATTTGAAAGACTACGCTGGCTTG	1355
Qy	201	GlyIleLeuHisAspIleGluGlyIleArgTyrGlnProLysCysProLeuGlyValAsp	220
Db	1356	GGCATTTTACATGACATGAAAGGCAATCCGCTATGAAACCAAAATGTCTCTTGGAGTAC	1415
Qy	221	IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle	239
Db	1416	ATATCAAAAGAGAGTTGGAGAAAGCTTCATTAAGTACCAACATTTAAATATGAGATA	1472

RESULT 7

US-10-651-722-1

/ Sequence 1, Application US/10651722

/ Publication No. US20040048302A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: Polynucleotide ENCODING A NOVEL METALLOPROTEASE, MP-1

FILE REFERENCE: D0073 DIV

CURRENT APPLICATION NUMBER: US/10/651,722

CURRENT FILING DATE: 2003-08-29

PRIOR APPLICATION NUMBER: US 60/266,518

PRIOR FILING DATE: 2001-02-05

PRIOR APPLICATION NUMBER: US 10/067,443

PRIOR FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/282,814

PRIOR FILING DATE: 2001-04-10

NUMBER OF SEQ ID NOS: 71

SOFTWARE: PatentIn version 3.2

SEQ ID NO 1

LENGTH: 2197

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

/ LOCATION: (231)..(1472)

US-10-651-722-1

Alignment Scores:

Pred. No.: 1,79e-151

Score: 1240.00

Length: 2197

Matches: 239

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 17 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-651-722-1 (1-2197)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyValSerLeuAspIle 20  
Db 756 CTGTTGGCATTGATGTTCAAGAGATTTCAGATTTTCGTCTTCTTGAAAGTCTTTGACATA 815  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db 816 GCACCAAGGTGACATGCTTGAACAGGTGGCAAGAGACTTTCTTTAATTAACATCCAGAG 875  
QY 41 CysSerThrMetSerGlyLysValAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
Db 876 TGCTCCACCATGATGTTGGGAAAGCCATAGAACATTGGCCCAAAACAGGAATATGATTT 935  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
Db 936 CATTGTGACATCAAACTCTCCCTTGATCATGCTCAAAATGTTGATTTTCTTTTACCTGGA 995  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 100  
Db 996 CTTCAACACGTTACTGATTAATTAATGAAAGAAAGAAAGAAAGGATATTGAGAG 1055  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisIsthMetAlaCys 120  
Db 1056 GGGCAAACTCTGTCTTCAACAGACAGACATTGCTGCCACATGACACACATGCGCATGT 1115  
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
Db 1116 CATCTTGAGAAAAGAACACATGCGGCTATTTCTTTTGTAAAGCAGACAGACTGTACT 1175  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheTyrlleArgArg 160  
Db 1176 CAATTAATAGCAGACTGTGATGATCTGGGTGGTGGCAAGTAATCTTAATCCGACGA 1235  
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
Db 1236 GCTCGGAAATTTTAAACAACGACACACAGTGCACTTGTGTGCTCTCCACAGCTA 1295  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuAlaGlyLeu 200  
Db 1296 TGCACGTGATTAAGGCAATTGATTCATGCAATGATTAAGAAACATACGTCGCTTG 1355  
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgTyrlleArgCysProLeuGlyValAsp 220  
Db 1356 GGCATTTTAAACAGACATTAAGAGCAATCCGCTAAGAACCAAAAGTCTCTTGAGTGAAC 1415  
QY 221 IleSerLysGlnValGlyGlnLysSerIleLysValProGlnLeuLysMetGluIle 239  
Db 1416 ATATCAAAAGAGCTTGAGAGACCTTCATTAAGTACCAATTAATAAATGAGATA 1472

RESULT 8  
US-10-067-443-21  
; Sequence 21, Application US/10067443  
; Publication No. US20030082782A1  
GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
; FILE REFERENCE: D0073 NP  
; CURRENT APPLICATION NUMBER: US/10/067,443  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US 60/266,518  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: US 60/292,914  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: Patented version 3.0  
; SEQ ID NO 21

LENGTH: 1387  
TYPE: DNA  
ORGANISM: homo sapiens  
US-10-067-443-21

## Alignment Scores:

Pred. No.:	2,72e-148	Length:	1387
Score:	1213.50	Matches:	238
Percent Similarity:	90.15%	Conservative:	0
Best Local Similarity:	90.15%	Mismatches:	1
Query Match:	97.86%	Indels:	25
DB:	14	Gaps:	1

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-067-443-21 (1-1387)

```
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 549 CTGTTGGCATTAGTTCAGAGAGTTTCAGATTTTCGCTTCTTGGAAGCTTTGGACATA 608
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 609 GCACCAAGGTGACATCTTGACAGAGTGGCAGAGACCTTTCTTTAATTAACATCCAGAG 668
QY 41 CysSerThrMetSerGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 669 TGCTCCACCATGAGTGGTGGGAAAGCCATAGAGCATTTGGCCAAACAAGAAATGATTT 728
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
DB 729 CATTTTGACATCAAACTCCCTGTCATCATGCTAAATAATGATTTTCTTTTACTGGA 788
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIle----- 98
DB 789 CTTCAACACGTTTACTGATTAATAATTAATGAAGAAAGAAAGAGAGATTAATTCTTA 848
QY 98 ----- 98
DB 849 ATTGTAAGTTGAACAGATAAATATTCCTGATTTGCTGCTAAATAAGTCTCATTTTC 908
QY 99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115
DB 909 TGCAGGTATGAGAGGGGCAAAATCCTGCTTTCAGACAGACATTTGCCACAGTACAG 968
QY 116 HisThrMetAlaCyHisIleuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135
DB 969 CACCAATGGCATGTCATCTTGTAAGAAAGAACACATCGGGCTAATCTGTTTGTAGCAG 1028
QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsn 155
DB 1029 AGAGACTTGTACCTCAAAATTAATGCAATCTGCTGATCTGCTGCTGCTGCTGCTGCT 1088
QY 156 PheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCy 175
DB 1089 TTCTATATCCGCAAGCTCTGGAATTTTAAACAACGCAACAGTGCATTTGTTGTGT 1148
QY 176 ProProProAlaGluLeuCyThrAspAsnGlyIleMetIleAlaIleThrAsnGlyIleGluArg 195
DB 1149 CCTCTCCCAACACTATGACATGATATGCAATTAATGATGATGATGATGATGATGATG 1208
QY 196 LeuArgAlaGlyLeuGlyIleLeuHisAspIleGluGlyIleArgTyGluProLysCys 215
DB 1209 CTAGCTGTGGCTTGCGGCAATTTTACATACATAGAAAGGATCGCTTAGAACCAAAATGT 1268
QY 216 ProLeuGlyValAspIleSerLysGlyValGlyGluAlaSerIleLysValProGlnLeu 235
DB 1269 CCTCTTGGAGTAGACATATCAAAAGAGTGGAGAGAGCTTCATTAAGTATCCACATTA 1328
QY 236 LysMetGluIle 239
DB 1329 AAAATGGAGATA 1340
RESULT 9
US-10-649-273-21
```

```
Sequence 21, Application US/10649273
Publication No. US20040043407A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
FILE REFERENCE: D0073 CNT
CURRENT APPLICATION NUMBER: US/10/649,273
PRIOR FILING DATE: 2003-08-27
PRIOR APPLICATION NUMBER: US 60/266,518
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 10/067,443
PRIOR FILING DATE: 2002-02-05
PRIOR APPLICATION NUMBER: US 60/282,814
PRIOR FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn version 3.2
SEQ ID NO 21
LENGTH: 1387
TYPE: DNA
ORGANISM: homo sapiens
US-10-649-273-21

Alignment Scores:
Pred. No.: 2,72e-148 Length: 1387
Score: 1213.50 Matches: 238
Percent Similarity: 90.15% Conservative: 0
Best Local Similarity: 90.15% Mismatches: 1
Query Match: 97.86% Indels: 25
DB: 17 Gaps: 1

US-10-649-273-2_COPY_176_414 (1-239) x US-10-649-273-21 (1-1387)
QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20
DB 549 CTGTTGGCATTAGTTCAGAGAGTTTCAGATTTTCGCTTCTTGGAAGCTTTGGACATA 608
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 609 GCACCAAGGTGACATCTTGACAGAGTGGCAGAGACCTTTCTTTAATTAACATCCAGAG 668
QY 41 CysSerThrMetSerGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 669 TGCTCCACCATGAGTGGTGGGAAAGCCATAGAGCATTTGGCCAAACAAGAAATGATTT 728
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCyAspPheSerPheThrGly 80
DB 729 CATTTTGACATCAAACTCCCTGTCATCATGCTAAATAATGATTTTCTTTTACTGGA 788
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIle----- 98
DB 789 CTTCAACACGTTTACTGATTAATAATTAATGAAGAAAGAAAGAGATTAATTCTTA 848
QY 98 ----- 98
DB 849 ATTGTAAGTTGAACAGATAAATATTCCTGATTTGCTGCTAAATAAGTCTCATTTTC 908
QY 99 -----GluLysGlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGln 115
DB 909 TGCAGGTATGAGAGGGGCAAAATCCTGCTTTCAGACAGACATTTGCCACAGTACAG 968
QY 116 HisThrMetAlaCyHisIleuValLysArgThrHisArgAlaIleLeuPheCysLysGln 135
DB 969 CACCAATGGCATGTCATCTTGTAAGAAAGAACACATCGGGCTAATCTGTTTGTAGCAG 1028
QY 136 ArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsn 155
DB 1029 AGAGACTTGTACCTCAAAATTAATGCAATCTGCTGATCTGCTGCTGCTGCTGCTGCT 1088
QY 156 PheTyrlleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCy 175
DB 1089 TTCTATATCCGCAAGCTCTGGAATTTTAAACAACGCAACAGTGCATTTGTTGTGT 1148
QY 176 ProProProAlaGluLeuCyThrAspAsnGlyIleMetIleAlaIleThrAsnGlyIleGluArg 195
```

[illegible]

**RESULT 10**

```

US-10-651-722-21
; Sequence 21, Application US/10651722
; Publication No. US20040048302A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE, MP-1
; FILE REFERENCE: D0073 DIV
; CURRENT APPLICATION NUMBER: US/10/651,722
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US 60/266,518
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 10/067,443
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/282,814
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-651-722-21

```

**Alignment Scores:**

Pred. No.:	2,72e-148	Length:	1187
Score:	1213.50	Matches:	238
Percent Similarity:	90.15%	Conservative:	0
Best Local Similarity:	90.15%	Mismatches:	1
Query Match:	97.86%	Indels:	25
DB:	17	Gaps:	1

US-10-649-2\_COPY\_176\_414 (1-239) x US-10-651-722-21 (1-1387)

Qy	1	Leu	Leu	Ala	Leu	Val	Gln	Gly	Val	Ser	Asp	Phe	Leu	Leu	Leu	Gln	Ily	Ser	Leu	Asp	Ile	20	
Db	549	C	T	G	T	T	G	C	A	T	T	G	A	G	A	T	T	T	C	T	T	G	608
Qy	21	Ala	Pro	Gly	Asp	Met	Leu	Asp	Leu	Val	Ala	Arg	Arg	Leu	Ser	Leu	Ile	Y	S	Pro	Glu	40	
Db	609	G	C	A	C	A	G	G	T	G	A	C	A	T	G	T	T	T	T	T	A	T	666
Qy	41	Cys	Ser	Thr	Met	Ser	Gly	Gly	Val	Ala	Ile	Gln	His	Leu	Ala	Ser	Gln	Ily	Asn	Arg	Phe	60	
Db	669	T	G	C	T	C	C	A	C	A	T	G	T	G	T	G	G	A	A	G	C	A	728
Qy	61	His	Phe	Asp	Ile	Leu	Pro	Pro	Leu	His	Ser	Ala	Ala	Val	Asn	Cys	Asp	Phe	Ser	Phe	Thr	80	
Db	729	C	A	T	T	T	G	A	C	A	T	C	A	A	A	C	T	T	C	C	T	G	788
Qy	81	Leu	Gln	His	Val	Thr	Asp	Leu	Ile	Ile	Met	Leu	Ser	Gln	Ily	Ser	Gln	Ily	Ile	-----	98		
Db	789	C	T	T	C	A	C	A	C	G	T	T	A	A	A	T	T	A	A	T	A	T	848
Qy	98	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	98		
Db	849	A	T	T	A	C	T	A	A	G	T	T	A	A	A	T	A	T	T	C	T	G	908

Oy	99	-----GluYSGIYGInIIeLeuseSerIalIAspRIeIalIathValGIIn	115
Dd	909	TGCAGATATGGAAAGGGCAAACTCTGTCTTACAGCAGACATTCCTCCACAGTAAC	968
Oy	116	HISrHMeTalAcYShIeLeuValLYaRqThriIsarGalaiIeLeuPheCYaLSeGIIn	135
Dd	969	CACACAATGGCATGTCAATCTTGTAAMAAAACAACATCGGGGTATTCTGGTTTGTAAACAG	1028
Oy	136	ArgAsPrleuPeProGInnaSPnaSnalIvalIeuValIaseRgIYGValAIaseRsn	155
Dd	1029	AGAAGCTGTGAACCTCAAAATAAAGCAGATCTGTGTGATCTGGTGGIGTCGCAAGTAAC	1088
Oy	156	PHeTYrIIeaRGaRGaIaleuGIuIIeueThzASnAlaThRngIcySthrLeuLeuCys	175
Dd	1089	TTCTATATCCGCAGAGCTCTGGAATTTTAAcAAACGACAACAGTGCATCTTGTGTGT	1148
Oy	176	ProPRoPRoMAdglueCySthrASpASnGIYLlEmetIlleaITPaNGIYLlEGLuaRQ	195
Dd	1149	CCCTCTCCCAAGACTATGACTGAATAATGGCACTTATGATGATGGAAATGGTATTTGAAGA	1208
Oy	196	LauARgaIagIyleuGIYLlEleuHIeASpRIeGLuGIYleatRYrgIuProlYSes	215
Dd	1209	CTAGCTGGTGCTTGGGCCATTTTACATGAGCATTAAGAAGCATCCGCTATGAACCMAANTGT	1268
Oy	216	ProLeuGIYValAspRIeSerLyseGIUaIGIYuLaaserIleLYsValProGInIleu	235
Dd	1269	CCCTCTTGAGTAGCATATCAAAAGAAGTGGAGAACCTTCCAAATAAAGTACCACAATTA	1328
Oy	236	LYseMeGIuIle 239	
Dd	1329	AAAAATGACGATA 1340	

## RESULT 11

US-10-012-140-6  
; Sequence 6, Application US/10012140  
; Publication No. US20030009017A1  
CENTRAL INFORMATION.

```

? GENERAL INFORMATION:
? APPLICANT: Leidy, Kevin R.
? APPLICANT: Kapeller-Libermann, Rosana
? APPLICANT: Glucksmann, Maria A.
? TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
? TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USSES
? TITLE OF INVENTION: THREEOP
? FILE REFERENCE: 381552004900
? CURRENT APPLICATION NUMBER: US/10/012,140
? CURRENT FILING DATE: 2001-11-08
? PRIOR APPLICATION NUMBER: 60/246,768
? PRIOR FILING DATE: 2000-11-08
? PRIOR APPLICATION NUMBER: 60/246,772
? PRIOR FILING DATE: 2000-11-08
? PRIOR APPLICATION NUMBER: 60/249,185
? PRIOR FILING DATE: 2000-11-15
? NUMBER OF SEQ ID NOS: 49
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 6
? LENGTH: 1245
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-10-012-140-6

```

### Alignment Scores:

Pred. No.:	5,52e-147	Length:	1245
Score:	1203.00	Matches:	232
Percent Similarity:	98.33%	Conservative:	3
Best Local Similarity:	97.07%	Mismatches:	4
Query Match:	97.02%	Indels:	0
DB:	14	Gaps:	0

US-10-649-273-2\_COPY\_176\_414 (1-229) x US-10-012-140-6 (1-1245)

Dy 1 LeulewalaLeuValGlnGlyValSerAspPheLeuLeuGlnIleYsSerLeuAspIle 20  
| | | | |  
Db 526 CTGTGGCAATTAGTCAAGAAGTTACAGATTTCTGCTTCTTGAAAGTCITTTGCACATA 585

```
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
D 586 GCACCAAGTGCATCTTGCACAGAGTGGCAAGAACCTTTCTTATATAACATCCAGG 645
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
D 646 TGCTCCACCATGAGTGGTGGAAAGCCATAGAACATTTGGCCAAACCAAGAAATAGATTT 705
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
D 706 CATTTTGACATCAACCTCCCTTCATCATCTGCTAAATAATGATTTCTTTTATCTGGA 765
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 100
D 766 CTTCACACGTTACTGATTAATAATAATGAACACAGAAACAAAGAGAGATTTAGAG 825
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
D 826 GGGCAAAATCCTGCTCTTCAGCAGCAGCATTTGCTGCACAGTACAGACACAAATGCATGT 885
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
D 886 CATCTTGGAAGAAAGACATCGGCTATCTGTTTGTAGCAGAGAGACTTGTTACCT 945
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
D 946 CAAATATAATGACGATCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
D 1006 GCTCGGAAATTTTAAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1065
QY 181 CysThrAspAsnGlyIleMetIleAlaArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
D 1066 TGCACTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 1125
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgGlyIleGluProLysCysProLeuGlyValAsp 220
D 1126 GGCATTTTACATGACATGAAAGGACATCCGCTATGAAACCAAAATGCTCTTGGAGTAGAC 1185
QY 221 IleSerLysGluValGlyGluLysSerIleLysValProGlnLeuLysMetGluIle 239
D 1186 ATATCAAAAGAGTGGAGAGCTTCCATTAAGTACCAATTAATAATGAGAGATA 1242

RESULT 12
US-10-012-140-4
; Sequence 4, Application US/10012140
; Publication No. US20030009017A1
; GENERAL INFORMATION:
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; TITLE OF INVENTION: 38650, 28472, 5495, 65507, 81588, AND
; TITLE OF INVENTION: 14354 METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USUS
; TITLE OF INVENTION: THEROOP
; FILE REFERENCE: 381552004900
; CURRENT APPLICATION NUMBER: US/10/012,140
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/246,768
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,772
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,185
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1820
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: (146) ... (1390)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) ... (1820)
; OTHER INFORMATION: n = A,T,C or G
US-10-012-140-4

Alignment Scores:
Pred. No.: 9 62e-147
Score: 1203.00
Percent Similarity: 98.33%
Best Local Similarity: 97.07%
Query Match: 97.02%
DB: 14
Gaps: 0

US-10-649-273-2_COPY_176_414 (1-239) x US-10-012-140-4 (1-1820)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnLysSerLeuAspIle 20
D 671 CTGTTGGCATTTAGTTCAAGAGATTTTCAAGATTTTCTGCTTCTGGAAGCTTTGGACATA 730
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
D 731 GCACCAAGTGCATGCTTGCACAGAGTGGCAAGAGACTTTCTTTAATAAACATCCAGAG 790
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
D 791 TGCTCCACCATGAGTGGTGGAAAGCCATAGAACATTTGGCCAAACAGGAATAGATTT 850
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
D 851 CATTTTGACATCAACCTCCCTTCATCATCTGCTAATAAATTTGATTTTCTTTTATCTGGA 910
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGlnLysGlnGlyIleGluLys 100
D 911 CTTCACACGTTACTGATTAATAATAATGAACACAGAAACAAAGAGAGATTTAGAGAG 970
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaCys 120
D 971 GGGCAAAATCCTGCTCTTCAGCAGCAGCATTTGCTGCACAGTACAGACACAAATGCATGT 1030
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140
D 1031 CATCTTGTAAGAAAGACATCGGCTATCTGTTTGTAGCAGAGAGACTTGTTACCT 1090
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheThrIleArgArg 160
D 1091 CAAATATAATGACGATCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1150
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProProArgLeu 180
D 1151 GCTCGGAAATTTTAAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1210
QY 181 CysThrAspAsnGlyIleMetIleAlaArgAsnGlyIleGluArgLeuArgAlaGlyLeu 200
D 1211 TGCACTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 1270
QY 201 GlyIleLeuHisAspIleGlnGlyIleArgGlyIleGluProLysCysProLeuGlyValAsp 220
D 1271 GGCATTTTACATGACATGAAAGGACATCCGCTATGAAACCAAAATGCTCTTGGAGTAGAC 1330
QY 221 IleSerLysGluValGlyGluLysSerIleLysValProGlnLeuLysMetGluIle 239
D 1331 ATATCAAAAGAGTGGAGAGCTTCCATTAAGTACCAATTAATAATGAGAGATA 1387

RESULT 13
US-10-094-749-400
; Sequence 400, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
```

```

APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKI
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
FILE REFERENCE: 08435/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 400
LENGTH: 2208
TYPE: DNA
ORGANISM: Homo sapiens
US-10-094-749-400

```

```

Alignment Scores:
Pred. No.: 9,58e-128 Length: 2208
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Indels: 24
DB: 17 Gaps: 1

```

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-094-749-400 (1-2208)

```

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAspIle 20
DB 869 CTGTTGGCAATTAGTTCAGAGAGTTTCAGATTTCCTTGGAAAGCTTTGGACATA 928
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 929 GCACCAAGTGACATCTTGACAAAGGCGCAAGAGCTTCTTAATAAATCCAGAG 988
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 989 TGCTTCACCATGACGTGTGGGAAAGCCATAGAACTTTGGCCAAACAGAAATATGATTT 1048
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 1049 CATTGTGACATCAACCTCCCTGCAATGCTTAATAATTTGATTTCTTTTACTGGA 1108
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlnGlyIleGluLys 100
DB 1109 CTTCAACACGTTACGATTAATAATTAATGAAGAAAAGAGAAAGAGATTTAGAGAG 1168
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaIleAsnValGlnHisThrMetAlaCys 120
DB 1169 GGGCAAAATCTGTCTTCAAGCAGACATTCCTGCACAGTACAGCACAATGCGATGT 1228
QY 121 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPro 140
DB 1229 CATCTTGAAAAGAACACATGGGCTATCTGTTTGTAGACAGAGACATTTGTAACCT 1288
QY 141 CysAsnAsnAlaValIleuValAlaSerGlyGlyValAlaSerAsnPheTyrlleArgArg 160
DB 1289 CAAATATATGCAATCTGCTTGCATCTGTGTGCTGCGAGTACTTCTGTATCCGACA 1348

```

```

QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180
DB 1349 GCTCTGGAATTTTAAACAACGACACAGTGACCTTTGTGTCTCTCCCAAGACTA 1408
QY 181 CysThrAspAsnGlyIleMetIleAlaATPAsnGlyIleGluArgLeuArgAlaGlyLeu 200
DB 1409 TGCATGATTAATGCGCATTAATGATTTGCA----- 1435
QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220
DB 1436 -----TGATGTCTCTTGGAGTATGAC 1456
QY 221 IleSerLysGluValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239
DB 1457 ATATCAAAAGAACTTGAGAGAGCTTCCATTAAGTACCAATTAATAAATGAGATA 1513

```

```

RESULT 14
US-10-723-860-7447
Sequence 7447, Application US/10723860
Publication No. US20040253606A1
GENERAL INFORMATION:
APPLICANT: Aziz, Nacasha
APPLICANT: Ginsburg, Wendy M.
APPLICANT: Zlotnik, Albert
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
FILE REFERENCE: 05882.0193.NPUS01
CURRENT APPLICATION NUMBER: US/10/723,860
CURRENT FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 7447
LENGTH: 2890
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc.feature
LOCATION: (646)..(657)
OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-7447

```

```

Alignment Scores:
Pred. No.: 1.42e-127 Length: 2890
Score: 1059.00 Matches: 211
Percent Similarity: 88.28% Conservative: 0
Best Local Similarity: 88.28% Mismatches: 4
Query Match: 85.40% Indels: 24
DB: 18 Gaps: 1

```

US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-723-860-7447 (1-2890)

```

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlnGlySerLeuAspIle 20
DB 1526 CTGTTGGCAATTAGTTCAGAGAGTTTCAGATTTCCTTGGAAAGCTTTGGACATA 1585
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40
DB 1586 GCACCAAGTGACATCTTGACAAAGGCGCAAGAGCTTCTTAATAAATCCAGAG 1645
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60
DB 1646 TGCTTCACCATGACGTGTGGGAAAGCCATAGAACTTTGGCCAAACAGAAATATGATTT 1705
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80
DB 1706 CATTGTGACATCAACCTCCCTGCAATGCTTAATAATTTGATTTCTTTTACTGGA 1765
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlnGlyIleGluLys 100
DB 1766 CTTCAACGTTTACTAATAATTAATGAAGAAAAGAGAGATTTAGAGAG 1825

```



Qy 101 GlyGlnIleLeuSerSerAlaIleAspIleAlaIleThrValGlnIleIleThrMetAlaCys 120  
Db 1826 GGGCAATCTCTCTTCAGCAGCAGACATCTGTCACAGTACAGCAGCAATGGCATGT 1885  
Qy 121 HisIleValIleAspThrHisArgAlaIleLeuPheCysIleGlnIleArgAspLeuPro 140  
Db 1886 CATCTGTGAAAAGAACATCGGGCTATTCTGTTTGTAGCAGAGACATTTGTACCT 1945  
Qy 141 GlnAsnAsnAlaValIleuValAlaSerGlyValAlaSerAsnPheTyrIleArgArg 160  
Db 1946 CAAATTAATGCAAGTCTGTCATCTGCTGTGTCGCAAGTAACCTTGTAATCCGCA 2005  
Qy 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
Db 2006 GCTCTGGAATTTTAACAACCAACGACAGTGCATTTGTGTCTCTCTCCAGACTA 2065  
Qy 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
Db 2066 TGCACTGATTAATGCATTAATGATGCA----- 2092  
Qy 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLysCysProLeuGlyValAsp 220  
Db 2093 -----TGATGCTCTCTTGAGATGAC 2113  
Qy 221 IleSerIleGluValGlyValAlaSerIleLeuValProGlnLeuIleMetGluIle 239  
Db 2114 ATATCAAAAGAAAGTTGGAGAGCTTCCATTAAGTACCAATTAAATGAGATTA 2170  
RESULT 15  
US-10-067-443-20  
: Sequence 20, Application US/10067443  
: Publication No. US20030082782A1  
GENERAL INFORMATION:  
: APPLICANT: Bristol-Myers Squibb Company  
: TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL METALLOPROTEASE HIGHLY EXPRESSED IN  
: FILE REFERENCE: D0073 NP  
: CURRENT APPLICATION NUMBER: US/10/067,443  
: PRIOR FILING DATE: 2002-02-05  
: PRIOR APPLICATION NUMBER: US 60/266,518  
: PRIOR FILING DATE: 2001-02-05  
: PRIOR APPLICATION NUMBER: US 60/282,814  
: NUMBER OF SEQ ID NOS: 71  
: SOFTWARE: Patencin version 3.0  
: SEQ ID NO 20  
: LENGTH: 14364  
: TYPE: DNA  
: ORGANISM: homo sapiens  
US-10-067-443-20  
Alignment Scores:  
Pred. No.: 8,936-83 Length: 14364  
Score: 725.00 Matches: 186  
Percent Similarity: 32.86% Conservative: 0  
Best Local Similarity: 32.86% Mismatches: 0  
Query Match: 58.47% Indels: 380  
DB: 14 Gaps: 3  
US-10-649-273-2\_COPY\_176\_414 (1-239) x US-10-067-443-20 (1-14364)  
Qy 29 ValAlaAlaGlyLeuSerIleuIleIleYshIleProGluCysSerThrMetSerGlyGlyIle 48  
Db 11840 GTGGCAAAAGACTTTCTTTTATTAACATCCAGAGTCTCCACCAATGAGTGTGGGAAA 11899  
Qy 49 AlaIleGlnIleLeuAlaIleGlnIleAsnArgPheHisPheAspIleIleYshProProLeu 68  
Db 11900 GCCATTAAGCAATTTGGCCAAAGAAATAGATTTCATTGTGACATCAACCTCCCTTG 11959  
Qy 69 HisHisAlaIleYshCysAspPheSerPheThrGlyLeuGlnHisValThrAspIle 88  
Db 11960 CATCATGTAAATAATGTGATTTTCTTTTACTGACATTCACAGCTTAATGATTAATA 12019

Qy 89 IleMetIleYshGlnIleGluGlu----- 96  
Db 12020 ATTAATGAAAAAGAAAAAGAGAGATATTTCTTAATTAGTAAGTGAACGATTAAT 12079  
Qy 97 -----GlyIleGlnIleGlnIle 103  
Db 12080 ATTCTGATTTGTGCTTAATAATAGCTGCTATTCTTGCAAGGATTAAGAAAGGGCAAT 12139  
Qy 103 IleuSerSerAlaIleAspIleAlaIleThrValGlnIleIleThrMetAlaCysHisIleuVal 123  
Db 12140 CCTGCTTCAGCAGCAGCAATTTGCTGCCACAGTACGACACAAATGCAATGCTATCTGTG 12199  
Qy 123 IlyAspArgThrHisArgAlaIleLeuPheCysIleGlnIleArgAspLeuProGlnIleAsn 143  
Db 12200 GAAAAAGAACATCGGGCTATTCTGTTTGTAAAGCAGAGAGACTTGTACCTCAAAATTA 12259  
Qy 143 nAlaValIleu----- 146  
Db 12260 TCAGATGACTGTAAGTTTATCTATTATAGTAATAGTACACTTGCAATATGTAC 12319  
Qy 146 ----- 146  
Db 12320 TTTTTCACAGACCTTGAACCTTGCTGTGATGATGAAAGACATCTTATGCTTATGCTAG 12379  
Qy 146 ----- 146  
Db 12380 CCTGACAGTATGAATTAATGACAGATAGAAAAGACTAAACGCCATTTCTGTACTAGTT 12439  
Qy 146 ----- 146  
Db 12440 TGGTAGCTTATAGGACAGCTGATATAGCTTATATGACATATAGTCTAATTTTGACTT 12499  
Qy 146 ----- 146  
Db 12500 CTGTGTGATTTAAAGAGGGCTTACATTAAGAAAGTAATGACAGTAACGTCTATCACT 12559  
Qy 146 ----- 146  
Db 12560 AATTTTAAAGAAATAGTGTGATTTCTTCACTCTTGATGAATCCCTTTGTTGTTGT 12619  
Qy 146 ----- 146  
Db 12620 TTTTAAATAGCCAGTCAATTTTATGACAGTGGAGGTGTATTCAACTTGTGTACACT 12679  
Qy 146 ----- 146  
Db 12680 AATGTATAAAGTCTGTATTAATCCATTAATTTGACAGCCAAATCCCTTAATATGTG 12739  
Qy 146 ----- 146  
Db 12740 CTTAAAGCCTTGACAAACATCCTGTTTAATCTGTATCTTAACCTTATTAATAAAT 12799  
Qy 146 ----- 146  
Db 12800 TATTAACATAAGTGGAAATGTTTAATGTAGTAATTCATATAGATGAATTTACATGG 12859  
Qy 146 ----- 146  
Db 12860 ATATCAAGAAATATTTTTCAGAGTATGTAGTAATAATGCAAAATATATAAATTTTC 12919  
Qy 146 ----- 146  
Db 12920 AGGCTTAATAATAGTGTACTATGATTAATAATTAATTAATAATTAATAGATGAAGGT 12979  
Qy 146 ----- 146  
Db 12980 TCGAAGAAATATATCAAAATGCTAGTAATGTTGTATGCTATTAAGATTTATAGTAAT 13039  
Qy 146 ----- 146  
Db 13040 TTTTCTTCAAAATTTTATATACATAGATATGATCTGCCATTAACCATCTCAAAAT 13099  
Qy 146 ----- 146

```

Db 13100 GGGATAGTTATTATTGTTAATGCTGATATTTTCTCAGGTTTAAATAGACGTTGGT 13159
QY 147 -----ValAlaSerGlyGI 151
Db 13160 TCATATCCATATNAGATAGTATTTTGGTTTCTCATTCCTTCAGGTGCACTGTGTGG 13219
QY 151 ValAlaSerAsnPhenylleArgArgAlaLeuGluileuThraAsnAlaThrGlnCy 171
Db 13220 TGTGGCAAGTAACTTCTATATCCGACAGCTCTGGAAATTTTAAACAACGCACACAGTG 13279
QY 171 sThrleuLeuCyProProArgLeuCyThrAspAsnGlyileMeCileAlaTrp-- 190
Db 13280 CACTTGTGTGTCTCTCCGACACTATGCACTGATATGCAATTATGCAATGCAATGTA 13339
QY 190 ----- 190
Db 13340 AGCCACAGATATACGTGCTTCACTCATTAATGTAATTAATGCAATTTATTCAT 13399
QY 190 ----- 190
Db 13400 ACTAAGCCTTCTCTCTGATCTTGAGCTATTTGATTTTAAATGCTTCTTATTTTA 13459
QY 191 --AsnGlyileGluArgLeuArgAlaGlyleuGlyileuHisAspIleGluGlyileA 210
Db 13460 GGAATGGTATTGAAGAAGCTACGTGCTGGCTTGGCATTTTACATGACATAGAAGCATCC 13519
QY 210 rGlyrGluProlys 214
Db 13520 GCTATGAACCAAG 13533

```

Search completed: February 17, 2005, 01:27:31  
 Job time : 500 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 12:38:06 ; Search time 10.9109 Seconds  
(without alignments)  
2107.605 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVQGVDFLLGKSLDI.....DISKEVGEASIKVQLKMEI 239

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	326	26.3	463	2	B84888
2	244	19.7	365	2	A82902
3	244	19.7	366	2	D97677
4	241.5	19.5	359	2	AB3274
5	238	19.2	387	2	E71711
6	237.5	19.2	367	2	F87257
7	231	18.6	341	2	H83572
8	230.5	18.6	344	2	E97707
9	224	18.1	342	2	H64074
10	221.5	17.9	323	2	G69388
11	213	17.2	337	2	AG0892
12	210.5	17.0	335	2	G70369
13	210	16.9	337	2	A10079
14	209	16.9	324	2	F75029
15	209	16.9	337	2	C91122
16	209	16.9	337	2	B85967
17	206	16.6	337	1	Q08CR6
18	205.5	16.6	348	2	D82807
19	204.5	16.5	534	2	H69056
20	202	16.3	324	2	C71215
21	201.5	16.2	421	2	T18825
22	201	16.2	354	2	C81040
23	200	16.1	325	2	A38108
24	199	16.0	354	2	C81986
25	195.5	15.8	340	2	B97011
26	195.5	15.8	346	2	F59766
27	195	15.7	353	2	T04567
28	195	15.7	412	2	T40899
29	193.5	15.6	327	2	G72411

30	193	15.6	348	2	S75548	staloglycoproteina
31	192	15.5	346	2	AF1820	staloglycoproteina
32	192	15.5	346	2	H70195	staloglycoproteina
33	189	15.2	336	2	C97888	O-sialoglycoprotei
34	188.5	15.2	344	2	AC1334	glycoprotein endop
35	186.5	15.0	344	2	H70737	glycoproteinase fa
36	186	15.0	336	2	B95015	glycoprotein endop
37	183	14.8	343	2	D83718	glycoproteinase fa
38	182	14.7	338	2	A71545	probable O-sialogl
39	180.5	14.6	539	2	A64441	O-sialoglycoprotei
40	180	14.5	246	2	G90187	O-sialoglycoprotei
41	179	14.4	344	2	AB1705	glycoprotein endop
42	178	14.4	336	2	B84936	O-sialoglycoprotei
43	174.5	14.1	344	2	H72106	O-sialoglycoprotei
44	174.5	14.1	344	2	B85105	O-sialoglycoprotei
45	174.5	14.1	374	2	T35581	probable O-sialogl

ALIGNMENTS

RESULT 1  
B84888  
probable O-sialoglycoprotein endopeptidase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Jun-2003  
C:Accession: B84888  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Xoo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: AB84420; MID:20083487; PMID:10617197  
A:Accession: B84888  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-463 <STO>  
A:Cross-references: GB:AF002093; NID:g2583127; PIDN:AB82636.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g45270  
A:Map position: 2  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match	26.3%	Score 326	DB 2	Length 463
Best Local Similarity	33.6%	Pred No. 4.3e-22		
Matches	82	Conservative 34	Mismatches 84	Indels 44
			Gaps	5
QY	1	LIALVQGVDFLLGKSLDIAPQMDKVARRLSLIKHPBCSTMSGKAIIEHLAKGNRF	60	
DB	221	LIALVQGVDFLLGKSLDIAPQMDKVARRLSLIKHPBCSTMSGKAIIEHLAKGNRF	275	
QY	61	HPDIKPLHAKNCDPSTGLQHTVDKIIIMKEKEBGEIKGQILSSADIAATVQHTMAC	120	
DB	276	SVKENVPKMYHDCNFSYAGLTKQVRLAIEAKG-----INRADIAASFQVAVL	325	
QY	121	HLVKTTHAILPCQKRDLLPQNNAVLVASGVASNPYIRALAILTNAOCTLLCPPRL	180	
DB	326	HLBEKCEBALDWALE---LEPSIKHNVISGVASNKYVRLINNIVENKRLKLVCPPESTL	382	
QY	181	CTDNGIMTAMNGIERLRAGILGILHDEIGIRYE-----PRCPGLGVDSK	224	
DB	383	CTDNGVMAVMTGLEHFRVG-----RYDEPPATEPDEYVDLPRWPLGSEBYAKG	432	
QY	225	VGEA 228		
DB	433	RSEA 436		

RESULT 2  
AB2902  
O-sialoglycoprotein endopeptidase gcp [imported] - Agrobacterium tumefaciens (strain C58  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C/Accession: AB2902  
R:Wood, D.W.; Seubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Moo, I.  
erage, G.; Gillet, M.; Grant, C.; Guenthermer, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, F.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, B.W.  
A>Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AB2902  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-365 <KUR>  
A:Cross-references: GB:AE006688; PIDN:AAL43632.1; PID:g17741154; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: gcp  
A:Map position: circular chromosome  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.7%; Score 244; DB 2; Length 365;  
Best Local Similarity 31.9%; Pred. No. 1.1e-14;  
Matches 73; Conservative 34; Mismatches 92; Indels 30; Gaps 7;

QY 2 LALVQGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIENHLAKGNRPH 61  
Db 149 LVLVAGVGSEYRWGTTIDALGEAFDKTAKLGL-PYP-----GGPAVENNAAGDDPR 201

QY 62 FDIKPEPLHNAKNCDFSTGL-----QHVTDKIIMKKEKEGIEKGQILSSADIAATVQHT 117  
Db 202 FPLPRPMVGEARLDPSFSGKLTAVQAATATAPLSEQD-----INDICASFQA 250

QY 118 MACHLVKTRTHAILFCRKORLLPQNN--VLVAGGVASNFYIRALBILTNATQCTLLC 175  
Db 251 VSRILKRIKRGIGLARFKVE--FPHINGEPALVAGVAVANGRIQTLQALCTHGFRAVA 308

QY 176 PPPLCTDNGIMIANNGIERLRAGLILHDIGIRYRCPGLGVDSKE 224  
Db 309 PPHRLCTDNAMIAMVAGLERMAG-----ROADALBVAFPSRWPLDGSAB 353

RESULT 3  
D97677  
probable o-sialoglycoprotein endopeptidase (glycoproteinase) [imported] - *Agrobacterium*  
C/Species: *Agrobacterium tumefaciens*  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C/Accession: D97677  
R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: D97677  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-366 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK88373.1; PID:g15157858; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR C 4806  
A:Map position: circular chromosome  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.7%; Score 244; DB 2; Length 366;  
Best Local Similarity 31.9%; Pred. No. 1.1e-14;  
Matches 73; Conservative 34; Mismatches 92; Indels 30; Gaps 7;

QY 2 LALVQGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIENHLAKGNRPH 61  
Db 150 LVLVAGVGSEYRWGTTIDALGEAFDKTAKLGL-PYP-----GGPAVENNAAGDDPR 202

QY 62 FDIKPEPLHNAKNCDFSTGL-----QHVTDKIIMKKEKEGIEKGQILSSADIAATVQHT 117  
Db 203 FPLPRPMVGEARLDPSFSGKLTAVQAATATAPLSEQD-----INDICASFQA 251

QY 118 MACHLVKTRTHAILFCRKORLLPQNN--VLVAGGVASNFYIRALBILTNATQCTLLC 175  
Db 252 VSRILKRIKRGIGLARFKVE--FPHINGEPALVAGVAVANGRIQTLQALCTHGFRAVA 309

QY 176 PPPLCTDNGIMIANNGIERLRAGLILHDIGIRYRCPGLGVDSKE 224  
Db 310 PPHRLCTDNAMIAMVAGLERMAG-----ROADALBVAFPSRWPLDGSAB 354

RESULT 4  
AB3274  
O-sialoglycoprotein endopeptidase (EC 3.4.24.57) [imported] - *Brucella melitensis* (strain  
C/Species: *Brucella melitensis*  
C/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
C/Accession: AB3274  
R:Delvecchio, V.G.; Kaprat, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Ios, T.; Ivanova, T.  
; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A>Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A:Reference number: AB3252; PMID:11756688  
A:Accession: AB3274  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-359 <KUR>  
A:Cross-references: GB:AE008917; PIDN:AAL51357.1; PID:g17982056; GSPDB:GN00190  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BME10175  
A:Map position: I  
C:Superfamily: O-sialoglycoprotein endopeptidase  
C:Keywords: hydrolase; metalloproteinase

Query Match 19.5%; Score 241.5; DB 2; Length 359;  
Best Local Similarity 34.5%; Pred. No. 1.9e-14;  
Matches 69; Conservative 27; Mismatches 77; Indels 27; Gaps 6;

QY 2 LALVQGVSDPFLILGKSLDIAPGDMLDKVARRLSLIKHPECSTMSGKAIENHLAKGNRPH 61  
Db 145 LVLVAGIGDYERLIGTTIDALGEAFDKTAKLGL-PYP-----GGPAVERMLQGDQKR 197

QY 62 FDIKPEPLHNAKNCDFSTGL-----QHVTDKIIMKKEKEGIEKGQILSSADIAATVQHT 116  
Db 198 FPLPRPMVGEARLDPSFSGKLTAVROTATELVPITDQ-----DVTIDICASFQA 245

QY 117 TWACHLVKTRTHAILFCRKORLLPQ--NNAVLVAGGVASNFYIRALBILTNATQCTLLC 175  
Db 246 AVADLTLSDRVGSLSERFKTE--FPDCATPSLVVAGVAVANKTLRALLENLCTHGFRAFTA 303

QY 176 PPPLCTDNGIMIANNGIER 195  
Db 304 PPLNCTDNAMIAMVAGAR 323

RESULT 5  
E71711  
probable o-sialoglycoprotein endopeptidase (gcp) RP037 - *Rickettsia prowazekii*  
C/Species: *Rickettsia prowazekii*  
C/Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 09-Jul-2004  
C/Accession: E71711  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A>Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499; PMID:9823893  
A:Accession: E71711  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-387 <AND>  
A:Cross-references: UNIPROT:Q9ZNA8; GB:AJ235270; GB:AJ235269; NID:93860572; PIDN:CAA14506  
A:Experimental source: strain Madrid B  
C:Genetics:  
A:Gene: gcp; RP037  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.2%; Score 238; DB 2; Length 387;  
Best Local Similarity 29.7%; Pred. No. 4.4e-14;  
Matches 70; Conservative 32; Mismatches 74; Indels 60; Gaps 6;

QY 5 VQGVSDFLILGKSLDIPAGMDLVKARRSLIKHPECSTWSGKAI EHLAKOGRNRFHDI 64  
DB 145 VLIGGKTKILGTTTIDDAAGBAFDKIAKSLGL-PPY-----GPRLEKLA VGGDPTRIAL 208  
QY 65 KPLIHHAKNCDPSFTGLQHTVDKIIMK-KEKEBEGIEKGQILSSADIAATVQHTMACHLV 123  
DB 198 PKPIINSGNMNSFGSLKTA VRTLMNLKEVNDV-----INDIA SFPTGAILIS 249  
QY 124 KRTTRAILFCQRLDIPONNAVIVASGVASNFYIRRALBITLNAOTCTLLCPPRLCTDNGMIAMNGIER 195  
DB 250 SKMGDAIRLVYQILNDYEDINHPKTNLKSFRDGFPMKPLECTTRPKRYRHIHONSTRS 309  
QY 143 ---NAVLVASGVASNFYIRRALBITLNAOTCTLLCPPRLCTDNGMIAMNGIER 195  
DB 310 NLNDITVIVAGVANAKYLGELISDCTRPYGRILAPPMHLCTDMAAMIAVAGLER 365

RESULT 6

Peptidase M22 family protein [imported] - Caulobacter crescentus  
F87257  
C/Species: Caulobacter crescentus  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C/Accession: F87257  
B/Name: W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
R.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A/Title: Complete Genome Sequence of Caulobacter crescentus.  
A/Reference number: A87249; MUID:21173698; PMID:11259647  
A/Accession: F87257  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-367 <STO>  
A/Cross-references: GB:AE005673; NID:913421168; PIDN:AAK22058.1; GSPDB:GN00148  
C/Genetics:  
A/Gene: CC0071  
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 19.2%; Score 237.5; DB 2; Length 367;  
Best Local Similarity 34.3%; Pred. No. 4.6e-14;  
Matches 73; Conservative 29; Mismatches 90; Indels 21; Gaps 6;

QY 5 VQGVSDFLILGKSLDIPAGMDLVKARRSLIKHPECSTWSGKAI EHLAKOGRNRFHDI 64  
DB 156 VSGVGACKRLGTTTIDDAAGBAFDKIAKSLGL-PPY-----GPRLEKLA VGGDPTRIAL 208  
QY 65 KPLIHHAKNCDPSFTGLQHTVDKIIMK-KEKEBEGIEKGQILSSADIAATVQHTMACHLV 124  
DB 209 PRALIGRKDCDPSFGSLKTAARLAEITLTTDD-----ARRDLAAGVQAAIARQISE 259  
QY 125 RTRRAILFCQRLDIPONNAVIVASGVASNFYIRRALBITLNAOTCTLLCPPRLCTDNGMIAMNGIER 184  
DB 260 RVDRAAMLYR-DSHDEPDLRFVAVAGVANAGVAAVALLADCEKRGSPAPPLAVCTDN 317  
QY 185 GIMIANNGIERLAGLILHDIEGIRYEPKPL 217  
DB 318 AAMIALAGARL--ALGIFDDLDAIA-RPRWPL 347

RESULT 7

O-sialoglycoprotein endopeptidase PA0580 [imported] - Pseudomonas aeruginosa (strain PAO  
H83572  
C/Species: Pseudomonas aeruginosa  
C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C/Accession: H83572  
R/Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warener, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Lim,  
.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000  
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog  
A/Reference number: A82950; MUID:20437337; PMID:10984043  
A/Accession: H83572  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-341 <STO>  
A/Cross-references: GB:AE004494; GB:AE004091; NID:9946446; PIDN:AA03969.1; GSPDB:GN001;  
A/Experimental source: strain PA01  
C/Genetics:  
A/Gene: gcp; PA0580  
C/Superfamily: O-sialoglycoprotein endopeptidase

Query Match 18.6%; Score 231; DB 2; Length 341;  
Best Local Similarity 33.8%; Pred. No. 1.6e-13;  
Matches 73; Conservative 34; Mismatches 87; Indels 22; Gaps 8;

QY 2 LALVQGVSDFLILGKSLDIPAGMDLVKARRSLIKHPECSTWSGKAI EHLAKOGRNRFH 61  
DB 142 LVRVQIGIRYQILGSEVDDAAGBAFDKIAKSLGL-GYP-----GPRLEKLAERGTGR 194  
QY 62 FDIKPLIHHAKNCDPSFTGLQHTVDKIIMK-KEKEBEGIEKGQILSSADIAATVQHTMACH 121  
DB 195 FVFPFPMTRPGIDFSGSLKTFITLN-TWQRCVAGDDBSQ--TRCDIALAFQTAVER 250  
QY 122 LVKRTTRAILFCQRLDIPONNAVIVASGVASNFYIRRALBITLNAOTCTLLCPPRLCT 181  
DB 251 LLIKCRRL--KQGL--KN--LVLAGSVASNAQLRSGLEKMGKQVYARPRFC 302  
QY 182 TNGMIAMNGIERLAGLILHDIEGIRYEPKPL 217  
DB 303 TNGMIAMIAAGCQRLLAG---QHDPALISVQRPWM 335

RESULT 8

O-sialoglycoprotein endopeptidase (BC 3.4.24.57) [imported] - Rickettsia conorii (strain  
E97707  
C/Species: Rickettsia conorii  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
C/Accession: E97707  
R/Ogata, H.; Audic, S.; Reneseo-Audiffren, P.; Fournier, P.B.; Barbe, V.; Samson, D.; Ro  
Science 293, 2093-2098, 2001  
A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A/Reference number: A97700; MUID:21442074; PMID:11557893  
A/Accession: E97707  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-344 <KUR>  
A/Cross-references: GB:AB006914; PIDN:AA02599.1; PID:915619097; GSPDB:GN00173  
C/Genetics:  
A/Gene: gcp  
C/Superfamily: O-sialoglycoprotein endopeptidase  
C/Keywords: hydrolase; metalloproteinase

Query Match 18.6%; Score 230.5; DB 2; Length 344;  
Best Local Similarity 33.2%; Pred. No. 1.9e-13;  
Matches 64; Conservative 34; Mismatches 76; Indels 19; Gaps 5;

QY 5 VQGVSDFLILGKSLDIPAGMDLVKARRSLIKHPECSTWSGKAI EHLAKOGRNRFHDI 64  
DB 145 VLIGGKTKILGTTTIDDAAGBAFDKIAKSLGL-----AFPGSPLEKRAKADPRKTYKF 197  
QY 65 KPLIHHAKNCDPSFTGLQHTVDKIIMK-KEKEBEGIEKGQILSSADIAATVQHTMACHLV 123  
DB 198 PKPIINSGNMNSFGSLKTA VRTLMNLKEINDV-----INDIA SFPTGAILIS 249  
QY 124 KRTTRAILFCQRLDIPONNAVIVASGVASNFYIRRALBITLNAOTCTLLCPPRLCT 182  
DB 250 SKMGDAIRLVYQILNDYEDINHPKTNLKSFRDGFPMKPLECTTRPKRYRHIHONSTRS 307  
QY 183 DNGMIAMNGIER 195  
DB 308 DNAAAMIAVAGLER 320

## RESULT 9

H64074  
O:siatologlycoprotein endopeptidase (EC 3.4.24.57) - Haemophilus influenzae (strain Rd KW2  
N:Alternate names: siatologlycoproteinase  
C:Species: Haemophilus influenzae  
C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Jul-2004  
C:Accession: H64074  
R:Reischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.  
; Gocayne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrman, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: H64074  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-342 <TIGR>  
A:Cross-references: UNIPROT:P43764; GB:U32735; GB:I42023; NID:g1573509; PIDN:AAC22187.1;  
C:Superfamily: O-siatologlycoprotein endopeptidase  
C:Keywords: hydrolase, metalloproteinase

Query Match 18.1%; Score 224; DB 2; Length 342;  
Best Local Similarity 33.5%; Pred. No. 7,3e-13;  
Matches 67; Conservative 28; Mismatches 83; Indels 22; Gaps 6;

QY 2 LALVGVSDPFLILGKSLDIPAGMDLVKVARSLIKHPECSTMSGKAIIEHLAKQGNRFH 61

DB 142 LVRVDGKGVKYEIVGESIDDAAGEAFDPTAKLLGL-DYP-----GGALLSLRAEKGTPIR 194

QY 62 FDIKPEPLHAKNCDPSFTGLQHVTDKIIMKKEKEG--IEKQIILSSADIDATVQHTMA 119

DB 195 FTPEPRPMTDRAGLPDFSGKLTFTFPAANTVNOAIKNEGELIEQ-----TRADIAVAFQDAIV 249

QY 120 CHLVKRTTRAILFPCQRDLFPONNAVIVASGVSANFTIRALITLTNAQTCLCPPR 179

DB 250 DTLA-----IKCK-RALKETGYRVLVIAGVSANKKLRITLAHLMONLGSEVYPPQ 301

QY 180 LCTDNGIMIAMNGIERLRAG 199

DB 302 FCTDNGAMIAVTFGLRLKQ 321

## RESULT 10

G69388  
O:siatologlycoprotein endopeptidase homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000

C:Accession: G69388  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997  
A:Authors: Usterbeck, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: G69388

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-323 <KLB>  
A:Cross-references: GB:AE001027; GB:AE000782; NID:g2689350; PIDN:AAB90129.1; PID:g264947  
C:Superfamily: O-siatologlycoprotein endopeptidase

Query Match 17.9%; Score 221.5; DB 2; Length 323;

Best Local Similarity 32.8%; Pred. No. 1.2e-12;  
Matches 62; Conservative 30; Mismatches 66; Indels 31; Gaps 8;

QY 11 FLILGKSLDIPAGMDLVKVARSLIKHPECSTMSGKAIIEHLAKQGNRFHFDIPPLH 70

DB 142 YRVGERTLIDIGNALDKLAKHML-KHP-----GPKIEBLAKGQKTH--LP--YV 190

QY 71 AKNCDPSFTGLQHVTDKIIMKKEKEGKQIILSSADIDATVQHTMACHLVKTRTAL 130

DB 191 VKGMDPSFGVMTAAQRLEP-----DSGVA-----MEVAFSFOSTAPALTEVERAL 238

QY 131 LFCQKRDLLPQNNNAVIVASGVSANFTIRALITLTNAQTCLCPPRICDNGIMIAM 190

DB 239 AV-----LDLAEVILLV--GGVAANRLQLQMLRIMCEDBGAKFYVPPKELAGDNGAMIAV 290

QY 191 NGIERLRAG 199

DB 291 TGLIWKYKHG 299

## RESULT 11

AG0892  
probable glycoproteinase [imported] - Salmonella enterica subsp. enterica serovar Typhi (st

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AG0892

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AG0892

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-337 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD07733.1; PID:g16504285; GSPDB:GN00176

C:Gene: STY3387

C:Superfamily: O-siatologlycoprotein endopeptidase

Query Match 17.2%; Score 213; DB 2; Length 337;

Best Local Similarity 30.6%; Pred. No. 7.4e-12;  
Matches 66; Conservative 32; Mismatches 92; Indels 26; Gaps 7;

QY 2 LALVGVSDPFLILGKSLDIPAGMDLVKVARSLIKHPECSTMSGKAIIEHLAKQGNRFH 61

DB 142 LLSVTGIGYELGESIDDAAGEAFDPTAKLLGL-DYP-----GGPMLSKMASQGTAR 194

QY 62 FDIKPEPLHAKNCDPSFTGLQHVTDKIIMKKEKEGKQIILSSADIDATVQHTMACH 121

DB 195 FVPEPRPMTDRAGLPDFSGKLTFTFPAANTIRNSGDDE-----QTRADIAVAFEDAVDT 246

QY 122 LVKRTTRAILFPCQRDLFPONNAVIVASGVSANFTIRALITLTNAQTCLCPPRIC 181

DB 247 L-----MIKKK-RALBSTGFKRLVAMAGVSANRLTRAKLLEMMQKRGEVYFARPERC 298

QY 182 TDNGIMIAMNGIERLRAGGILHDIEGIVYERKCP 217

DB 299 TDNGAMIAVAGVRFKA--GVTADL-GVTVRPWP 331

## RESULT 12

G70369  
siatologlycoproteinase - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004

C:Accession: G70369

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove

V. Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: G70369

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-335 <AQF>

A:Cross-references: UNIPROT:O66986; GB:AB00708; NID:g2983356; PIDN:AAC06951.1; PID:g298  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 17.0%; Score 210.5; DB 2; Length 335;  
Best Local Similarity 32.0%; Pred. No. 1.2e-11;  
Matches 70; Conservative 41; Mismatches 73; Indels 35; Gaps 10;

QY 2 LALVGVSDPFLILGKSLDIAPGDMLDKVARSLIKHPECSTMSGKAIENHLAKGNRFH 61  
DB LLYVRDGRVDFLGGTLDLDAVGEAYDKVAKMGL-GYP-----GGPIIDRLAKGCKL- 192  
QY 62 FDIKRPPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 121  
DB 193 YPLPKPLMERGNLNFSSGLK---TALMLKKEKNVK-----EDIAVSQETVEI 242  
QY 122 LVKRTTRAILFCRKQDILLPQNNAVLVASGVASNFYIRALIEILTNAQ---CTLLCPP 178  
DB 243 LLEKS---LWAKKTKGIR---LVVVGVSANSRLR---FVFKKASQEVGFELIYHP 291  
QY 179 RLCTDNGIMTAMNGIERLRAGLGILHDIEGIRYEPKPL 217  
DB 292 SLSTDNMLMTIAYAGMERFKGVAVPLDVNP---QPNIP 327

## RESULT 13

A10079  
probable glycoproteinase gcp [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001

C:Accession: A10079  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: A10079  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC89500.1; PID:g15978736; GSPDB:GN00175  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 210; DB 2; Length 337;  
Best Local Similarity 30.0%; Pred. No. 1.4e-11;  
Matches 65; Conservative 33; Mismatches 91; Indels 28; Gaps 6;

QY 2 LALVGVSDPFLILGKSLDIAPGDMLDKVARSLIKHPECSTMSGKAIENHLAKGNRFH 61  
DB 142 LISVYGIGVYLLGESVDDAGEAFDKTKLGL-DYF-----GSPMLSRMAQGTGR 194  
QY 62 FDIKRPPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 121  
DB 195 FTFPRPMTDRBGDLDFSSGLKTPAANTIRANGDD-----QTRADIAFAFEDAVDT 246  
QY 122 LVKRTTRAILFCRKQDILLPQNNAVLVASGVASNFYIRALIEILTNAQCTLLCPPRL 180  
DB 247 LAKSKSA-----LDQGFRLVYAGVSANQTRLRKLADMMQRGSEVFPARPEF 297  
QY 181 CTDNGIMTAMNGIERLRAGLGILHDIEGIRYEPKPL 217  
DB 298 CTDNGIMTAMTAYAGVRLKSN---LASELSVSVPRPPL 331

## RESULT 14

F75029  
O-sialoglycoprotein endopeptidase (gcp) PAB1159 - Pyrococcus abyssi (strain Orray)  
C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C:Accession: F75029  
R:Anonymous, Genome  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc  
A:Reference number: A75001  
A:Accession: F75029  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-324 <KAW>  
A:Cross-references: UNIPROT:Q9UK77; GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50676  
A:Experimental source: strain Orray  
C:Genetics:  
A:Gene: gcp  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 209; DB 2; Length 324;  
Best Local Similarity 32.5%; Pred. No. 1.6e-11;  
Matches 65; Conservative 34; Mismatches 69; Indels 32; Gaps 8;

QY 1 LALVGVSDPFLILGKSLDIAPGDMLDKVARSLIKHPECSTMSGKAIENHLAKGNRF 60  
DB 135 VIALREGG--RYRVFGETYDIGNADIVFARIEGL-----GPGGPVKEKLAERGER 185  
QY 61 HDIKRPPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 120  
DB 186 ---IELP-YAVGMDLSEGL--LTEAIRKYSKGRV-----DLAYSFOETARA 230  
QY 121 HLKRTTRAILFCRKQDILLPQNNAVLVASGVASNFYIRALIEILTNAQCTLLCPPRL 180  
DB 231 ALVEYTERAVATE-----KDEVVLV--GVAANRLRLEMLRIMTEDEGKFPVPPYL 282  
QY 181 CTDNGIMTAMNGIERLRAGL 200  
DB 283 CRDNGIMTAYTGLRMVKAIGI 302

## RESULT 15

C91122  
probable O-sialoglycoprotein endopeptidase [imported] - Escherichia coli (strain O157:H7,  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C:Accession: C91122  
R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shindagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C91122  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-337 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA37370.1; PID:g13363420; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: EC83947  
C:Superfamily: O-sialoglycoprotein endopeptidase

Query Match 16.9%; Score 209; DB 2; Length 337;  
Best Local Similarity 30.6%; Pred. No. 1.7e-11;  
Matches 66; Conservative 34; Mismatches 90; Indels 26; Gaps 8;

QY 2 LALVGVSDPFLILGKSLDIAPGDMLDKVARSLIKHPECSTMSGKAIENHLAKGNRFH 61  
DB 142 LISVYGIGVYLLGESVDDAGEAFDKTKLGL-DYF-----GSPMLSRMAQGTGR 194  
QY 62 FDIKRPPLHAKNCDPSFTGLQHTVDKTIIMKKEKEGIEKGQILSSAADIATVQHTMAC 121  
DB 195 FTFPRPMTDRBGDLDFSSGLKTPAANTIRANGDD-----QTRADIAFAFEDAVDT 246  
QY 122 LVKRTTRAILFCRKQDILLPQNNAVLVASGVASNFYIRALIEILTNAQCTLLCPPRLC 181  
DB 247 LMKCKRAL-----DLGFKR--LVWAGVSANRLRLAKLAKMKKRGSEVFPARPEF 298

QY 182 TDNGIMAMNGIRLRAGAGILHIDIGIRYEPKCP 217  
||| ||| : : : : :  
Db 299 TDNGAMTAYAGMRFKA--GATADL-GVSYPKRWPL 331

Search completed: February 16, 2005, 13:09:23  
Job time : 11.9109 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 13:19:07 ; Search time 2839.68 Seconds  
(without alignments)  
3303.653 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 L1ALVGVSDPFLIKSLDI.....DISKVGSAIKVPQLKMEI 239

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues  
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh  
-Q/cgnt\_1/USPTO.spool/US10649273/runat\_14022005\_114703\_16411/app\_query.fasta\_1.1429  
-DB=EST -QPMT=fastcap -SUFFIX=rcst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=humand40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPMT=pro -NOR=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10649273.QCGN\_1\_1\_6799\_@runat\_14022005\_114703\_16411 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1118	90.2	870	5	BQ423651 AGENCOURT
2	1097	88.5	640	5	BQ636028
3	1079	87.0	2288	3	AK045669
4	1071	86.4	1622	3	AK011265
5	930	75.0	852	5	BX391919
6	926	74.7	1603	2	BC036671
7	908	73.2	701	2	BE740611
8	876	70.6	658	7	CP362328
9	838	67.6	637	7	CK941819

10	799	64.4	922	5	BQ961028
c 11	789	63.6	545	1	AV602901
12	784	63.2	490	6	CB852881
13	783	63.1	822	7	CF257246
14	753	60.7	866	5	BU127463
15	749	60.4	736	5	BU261251
16	746	60.2	723	5	BU043563
17	731.5	59.0	792	5	AM601179
18	717	57.8	634	2	CA057753
c 19	708	57.1	736	6	BU621780
c 20	707.5	57.1	701	5	BU032606
21	698	56.3	696	5	CB272391
22	685	55.2	597	6	CB72391
c 23	683	55.1	909	5	BX756548
c 24	682	55.0	548	7	CO880741
25	671	54.1	682	2	BM043703
26	671	54.1	730	7	CN823245
27	668	53.9	706	5	BU202465
c 28	666	53.7	869	5	BX754527
c 29	647	52.2	919	7	CF407294
c 30	646	52.1	484	1	AJ670918
c 31	646	52.1	878	5	BX776940
32	642	51.8	1173	6	CD508917
c 33	638	51.5	1082	5	BX359023
c 34	625	50.4	506	2	BF415802
35	614	49.5	861	5	BU246158
36	611	49.3	1171	5	BU261605
37	604	48.7	424	1	AA273921
38	600	48.4	863	5	BU376295
39	594	47.9	389	4	BM744822
40	591	47.7	812	5	BU246489
41	589	47.5	357	5	BO672554
42	581	46.9	704	5	BU272784
c 43	576	46.5	826	7	CR444994
c 44	573	46.2	385	1	AJ647827
45	572.5	46.2	613	1	AA920105

ALIGNMENTS

RESULT 1  
BQ423651  
LOCUS  
DEFINITION AGENCOURT\_7790948 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6065828  
5', mRNA sequence.  
ACCESSION BQ423651  
VERSION BQ423651.1 GI:21118966  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 870)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC/DC/DTP  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM13342 row: 1 column: 21  
High quality sequence contig: 710.  
Location/Qualifiers  
1..870  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6065828"

FEATURES

source

/tissue type="melanotic melanoma"  
/lab host="DH10B (phage-resistant)"  
/clone lib="NIH MGC 72"  
/note="Organ: skin; Vector: pCMV-SORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

## ORIGIN

## Alignment Scores:

Pred. No.: 3e-119 Length: 870  
Score: 1118.00 Matches: 220  
Percent Similarity: 98.22% Conservative: 1  
Best Local Similarity: 97.78% Mismatches: 2  
Query Match: 90.16% Indels: 2  
DB: 5 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BQ423651 (1-870)

QY 1 leuenuAlaLeuValGInGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
DB 196 CTGTTGGATTAGTTCAGAGAGTTTCAATTTCTGCTTGGAAAGCTTTGGACATA 255  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
DB 256 GCACCAAGTGACATGCTTGGACAGGTGGCAAGAGACCTTCTTATATAACATCCAGG 315  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
DB 316 TGCTCCACCATGAGTGTGGGAGAAAGCCATAGAACATTTGGCCAAACAGAAATAGATT 375  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
DB 376 CATTTGACATCAAACTCCCTTGATCATGCTAAATAATGTGATTTCTTTACTGGA 435  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysLysGluLysGluGlyIleGluLys 100  
DB 436 CTTCAACACGTTACTGATATAATAATATGAAAAAGAAAAAGAGAAAGTTTGAAG 495  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
DB 496 GGGCAATTCCTGCTTCAGACAGACATGCTGCCACAGTACACACATGTCATGT 555  
QY 121 HisIleuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
DB 556 CATCTTGTAAGAAACACATCGGCTATTCGTTTGTAAACAGACATGTTACT 615  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPheThrIleArgArg 160  
DB 616 CAAAATATATGACATGCTGCTGTCATCTGCTGTCGACAGTACTTCTATATCCGACGA 675  
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrIleLeuCysProProArgLeu 180  
DB 676 GCTCTGGAATTTTAAACAAACGCAACACAGTGCACCTTGGTGTCCTCTCCGACATA 735  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
DB 736 TGCACTATATGCAATTTATGATGATGCAATGATGATGATGATGATGATGATGATG 795  
QY 201 GlyIleLeuHisAspIleGluGlyIleArgGlyArgu-ProLysCysPro-LeuGlyValA 220  
DB 796 GGCATTTTACATGACATGAAAGCATCCGCTATGAAACCAATGATGCTCTTGGAGTGG 855  
QY 220 spIleSerLys 223  
DB 856 ACATATCAAAA 866  
RESULT 2  
BQ636028  
LOCUS BQ636028 640 bp mRNA linear EST 15-JUL-2002  
DEFINITION hd03d11.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he  
Homo sapiens cDNA clone hd03d11 5', mRNA sequence.  
ACCESSION BQ636028

## VERSION

BQ636028.1 GI:21760487  
EST.

## KEYWORDS

Homo sapiens (human)

## SOURCE

Homo sapiens

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

1 (bases 1 to 640)  
Wistow, G., Bernstein, S.L., Wyatt, M.K., Ray, S., Behal, A.,  
Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K.

## AUTHORS

Expressed sequence tag analysis of human retina for the NRI Bank  
Project: Retbindin, an abundant, novel retinal cDNA and alternative  
splicing of other retina-preferred gene transcripts

## TITLE

Mol. Vis. 8 (4), 196-204 (2002)

## JOURNAL

22103461

## MEDLINE

12107411

## PUBMED

Contact: Wistow G  
Section on Molecular Structure and Function  
National Eye Institute  
6/331, NIH, Bethesda, MD 20892-2740, USA  
Tel: 301 402 3452  
Fax: 301 496 0078  
Email: graeme@helix.nih.gov  
Plate: 03 row: d column: 11  
Seq primer: M13Rpl reverse primer (ABI).  
Location/Qualifiers

## FEATURES

source

1..640

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="hd03d11"

/tissue\_type="Retina"

/dev\_stage="Adult"

/lab\_host="EMDH10B"

/clone\_lib="Human Retina cDNA (Un-normalized,  
unamplified): hd/he"

/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue  
was dissected from two 80 year old donors with no observed  
eye disease. 100ug of total RNA was used for library  
construction. A directionally cloned cDNA library in the  
pSPORT1 vector (Life Technologies) was constructed at  
Bioserve Biotechnology (Laurel MD) essentially following  
the protocols of the SuperScript Plasmid System full  
details of which are contained in the manufacturer's  
instruction manual (http://www.lifetech.com/). First  
strand synthesis was carried out using a Not I  
primer-adaptor  
[5'-GGACATGTTCTAGATCGGAGCGCCGCT(T)15-3']. EST analysis  
was performed on the unamplified library at the NIH  
Intramural Sequencing Center (NISC)."

## ORIGIN

## Alignment Scores:

Pred. No.: 5.41e-117 Length: 640  
Score: 1097.00 Matches: 209  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 88.47% Indels: 0  
DB: 5 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BQ636028 (1-640)

QY 31 ArgArgLeuSerLeuIleLysHisProGluCysSerThrMetSerGlyGlyLysAlaIle 50  
DB 3 AGAAGACCTTCTTATATAAACAATCCAGAGGTGCTCCACATGATGATGATGATGATGATG 62  
QY 51 GlnHisLeuAlaLysGlnGlyAsnArgPheHisPheAspIleLysProProLeuHisHis 70  
DB 63 GAACATTTGGCCAAACAGAAATATATTTCAATTTGACATCAACCTCCCTTGCATCAT 122  
QY 71 AlaLysAsnCysAspPheSerPheThrGlyLeuGlnHisValThrAspLysIleIleMet 90  
DB 123 GCTAATAAATATGATTTTCTTTTACTGACATTCACAGCTTACTGATATAATATATATG 182

QY 91 Lysylsglytysglutylglytyleuylglytyleuylserinylalaanaspile 110  
 Db 183 AAAAAGAGAAAAAGAGAAAGATGAGAAAGGGCAAAATCGTCTTCCAGCAGACAGATT 242  
 QY 111 Alalaathrvalaiglnhiethrmetalaacyethislenvalylargthrhiatrgalalle 130  
 Db 243 GCTGCCACAGTACAGACACACATGCGATGCTCTTGAAAAAGAACATCGGGCTATT 302  
 QY 131 Leuphēcysylsrglnaraspaleupproglinaenanaalavalienvalalasergly 150  
 Db 303 CTGTTTGTAGACAGACAGACTTGTACTCAAAATAATGCAATCTGGTTCATCTGAT 362  
 QY 151 GlyvalalaseranpheTyrlleargrgalaleuglnulleuylthrAsnalathrgln 170  
 Db 363 GGTTCCCAAGTAATCTTATATCGCAGAGCTGTGAATTTTAACAACGCAACAG 422  
 QY 171 Cyethreuleucyepropropargleucyethrappanglyllemetllealatr 190  
 Db 423 TGCACTTGTGTGTCTCTCCCTCCAGACTATGCACTGATATGGCATTATGATTCATGG 482  
 QY 191 Asnglylleleuargleuargalaglyleuylleleuylasppilleglutyllearg 210  
 Db 483 AATGCTTTTGAAGAACTACGCTGCTGGCTTTCATGACATGAGAGCAATCCGC 542  
 QY 211 Tytgleubrolyscysproleuglyvalasppilesertysgluvalglylualaserlle 230  
 Db 543 TATGAACCAAAATGTCCTCTTGAGTAGACATATCAAAAGAAAGTTGAGAACCTTCATTA 602  
 QY 231 LysVal-ProglnLeuLysMetGluIle 239  
 Db 603 AAAGTACCACAAATTAATAATGAGATA 629

RESULT 3  
 LOCUS AK045669 2284 bp mRNA linear HTC 03-APR-2004  
 DEFINITION Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:B230219017 product:similar to PUTATIVE SIMLOGLYCOPROTEASE TYPE 2 [Homo sapiens], full insert sequence.  
 ACCESSION AK045669  
 VERSION AK045669.1 GI:26337528  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)  
 JOURNAL 99279253  
 MEDLINE 10349636  
 PUBMED 2

REFERENCE  
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)  
 JOURNAL 20499374  
 MEDLINE 11042159  
 PUBMED 3

REFERENCE  
 3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, O., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer Genome Res. 10 (11), 1757-1771 (2000)  
 JOURNAL 20530913  
 MEDLINE 11076861  
 PUBMED

REFERENCE  
 4 The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)  
 JOURNAL 5  
 REFERENCE  
 5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)  
 JOURNAL 6 (bases 1 to 2284)  
 REFERENCE  
 6 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, Y., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsumi, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submision

## COMMENT

Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/  
 URL: http://fantom.gsc.riken.jp/.

## FEATURES

## source

1..2284

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM\_DB:B230219017"

/db\_xref="taxon:10090"

/clone="B230219017"

/sex="male"

/tissue\_type="corpora quadrigemina"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

221..1465

/note="unnamed protein product; putative similar to PUTATIVE SIMLOGLYCOPROTEASE TYPE 2 [Homo sapiens] (SPTRLQ9H4B0, evidence: PASTY, 80%id, 100%length, match=142)"

/codon\_start=1

/protein\_id="BAC32450.1"

/db\_xref="GI:26337529"

/translation="MIMGRRTGALPKPKSKVYGLRPSVHPRTISCHKLVIGIET

SCDDTGAIVDRTGVNGLGALHSQTVHAKTGCIYPPVAQQLHRENIQIVERTLSAC

RITPDSIAIATITIRGLALSGVLSPLQVNOFPKPTIIRHEAALITRLTNK

VBPPLVLLISGCGLLLVQGVDFLLGSLDIAPGMDLRVARSLIRHPCST

MSGCAIQLADKGRFPTIIPPNQANCGPFGQIHITDKLITHEKEGLEKG

OLISGAADIIAAVQHTAHLAKRTHRALIPKQKNLS PANAIVVSGVASNLYIR

KALERVAANTCTLCPPPRCTDNGINIMANGIERLRAGLVLDVEDIRERPCPL

GVDSIRVAEALIKVPRLEKML"

2263..2267

/note="putative"

2284

/note="putative"

polyA\_site

polyA\_signal

## ORIGIN





```

|||||
Db 542 CTGTTGGCATTAGTTCAAGAGATTTCAGATTTTCGCTTCTTGGAAGCTTTGGACATA 483
Qy 21 A1AATPGIYASPMETLEUAPLYSVALA1AARGAGLEUSETLEU1ELYSH1SPROGLU 40
Db 482 GCACCAAGGTGACATGCTTGACAGAGGTGCAGAGACATTTCTTTATATAAACATCCAGAG 423
Qy 41 CysSerThmTserSergLYLYSVAL1EG1UH1SLEUALYSG1UGLYVANAARGPHE 60
Db 422 TGCTCCACCATGATGATGTTGGAAAGCCATAGAACATTTGGCCAAACAGAAATTCATTT 363
Qy 61 HisPheAP11elysProProLeuH1SH1A1ALYSAANCYASPPheserPheThngly 80
Db 362 CATTTCGATCAATCAACCTCCCTGTCATCATGTAAATTTGATTTTCTTTACTGCA 303
Qy 81 LeuGlnH1SVAlThraSPlyS11e11emelysLYSG1UGLYUGLYUGLYUGLYUGLY 100
Db 302 CTTCAACACGTTACTGATTAATATATATATATATATATATATATATATATATATATAT 243
Qy 101 G1YGIN111e1euserSerSerg1A1AAsp111e1A1a1ThraVAlGlnH1SthmTclAcyS 120
Db 242 GGGCAATCTCTGCTTTCAGACAGACATTCCTGCGACAGTACAGACACAAATGGCATGT 183
Qy 121 HisLeuVallySArgThrHisArgN1a11e1eupheCysLYSG1NARGSPheLeuPro 140
Db 182 CATCTTGGAAGAAAGACACATCGGCTATTCTGTTTGTAAACAGACAGACATGTTACT 123
Qy 141 G1AABNAN1A1AValLeuVAlA1AserGLYLYSVAL1A1AserAPNPhetYr11eArgArg 160
Db 122 CAAATATATGACATGCTGTCATCTGCTGTCGACAGTAACTTCTATATCCGACAGA 63
Qy 161 A1A1eug1u11e1e1ThraSP1A1a1ThrG1NCysThr11e1e1e1e1e1e1e1e1e1e1e 180
Db 62 GCTCTGGAATTTTAAACACACACACACACAGTGCATCTGTGTGTCTCTCCACAGACTA 3

```

RESULT 6  
 BC030671  
 LOCUS 1609 bp mRNA linear HTC 19-NOV-2003  
 DEFINITION Mus musculus O-sialoglycoprotein endopeptidase-like 1, mRNA (CDNA  
 clone IMAGE:1226118), containing frame-shift errors.  
 ACCESSION BC030671  
 VERSION BC030671.1 GI:21040459  
 KEYWORDS HTC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS  
 Straubeberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Sherman, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Wax, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Steplacet, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Schaefer, T.E., Brownstein, M.J., Ustin, T.B., Toshitaki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 McEwan, K.C., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Woleley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Vallalath, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettelman, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Sanchez, J., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bonfield, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Gilmwood, J., Schmitz, J., Myers, R.M.,  
 Butcherfield, Y.S., Krzywinski, M.I., Skalska, U., Smallue, D.E.,  
 Scherch, A., Schein, J.B., Jones, S.J., Jones, S.J., Jones, S.J., Jones, S.J.,  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 JOURNAL MEDLINE  
 PUBMED 12477932  
 REFERENCE 2 (bases 1 to 1609)  
 AUTHORS Straubeberg, R.

TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (20-MAY-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Marcello Bento Soares, Ph.D.  
 CDNA Library Preparation: M. Bento Soares, University of Iowa  
 DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
 contact: [amadansystemsbiology.org](mailto:amadansystemsbiology.org)  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha  
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

FEATURES  
 source  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRAP Plate: 66 Row: e Column: 10  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 21312463  
 This clone has the following problem: frame shifted.

Location/Qualifiers  
 1..1609  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1226118"  
 /tissue\_type="Thymus gland, mouse"  
 /clone\_lib="Soares\_thymus\_2NMNT"  
 /lab\_host="DH10B"  
 /note="Vector: pTTT3-Pac"

## ORIGIN

## Alignment Scores:

Pred. No.: 1,659-96 Length: 1609  
 Score: 926.00 Matches: 183  
 Percent Similarity: 81.59% Conservative: 12  
 Best Local Similarity: 76.57% Mismatches: 20  
 Query Match: 74.68% Indels: 24  
 DB: 3 Gaps: 1

US-10-649-273-2\_COPY\_176\_414 (1-239) x BC030671 (1-1609)

```

Qy 1 LeuLeuAlaLeuValGlnGlyValSerAPheLeuLeuGlyLSerSerLeuAps11e 20
Db 749 CTGTTGGCATTAGTTCAAGAGTTGTTCCGATTTCTCTCTTGGAAGCTTTGGACATA 808
Qy 21 A1AATPGIYASPMETLEUAPLYSVALA1AARGAGLEUSETLEU1ELYSH1SPROGLU 40
Db 809 GCACCAAGGTGACATGCTTGACAGAGGTGCAGAGAGACTTTCTTTATATCAACATCCAGAA 868
Qy 41 CysSerThmTserSergLYLYSVAL1EG1UH1SLEUALYSG1UGLYUGLYUGLYUGLY 60
Db 869 TGCTCCACCATGATGATGTTGGAAAGCCATAGAACATTTGGCCAAACAGAAATTCATTT 928
Qy 61 HisPheAP11elysProProLeuH1SH1A1ALYSAANCYASPPheserPheThngly 80
Db 929 CATTTCGATCAATCAACCTCCCTGTCATCATGTAAATTTGATTTTCTTTACTGCA 988
Qy 81 LeuGlnH1SVAlThraSPlyS11e11emelysLYSG1UGLYUGLYUGLYUGLYUGLY 100
Db 989 CTTCAACATATTAATCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1048
Qy 101 G1YGIN111e1euserSerSerg1A1AAsp111e1A1a1ThraVAlGlnH1SthmTclAcyS 120
Db 1049 GGGCAATCTCTGCTTTCAGACAGACATTCCTGCGACAGTACAGACATGCAACACGCTGC 1108
Qy 121 HisLeuVallySArgThrHisArgN1a11e1eupheCysLYSG1NARGSPheLeuPro 140
Db 1109 CACCTTGGAAGAAAGACACATCGGCTATTCTGTTTGGACAGAGAAATTTGCTCTCT 1168

```

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyValAlaSerAsnPhenylleArgArg 160  
 Db 1169 CAGCTAACGAGATTAAGTTAGTATCTGGAGGTGTGCAAGTAATTGACATCGGAAA 1228  
 QY 161 AlaleuGluLeuLeuThraAsnAlaThrGlnCysThrLeuLeuCysProProAlaArgLeu 180  
 Db 1229 GCATTGGAAATATGCGCAAAATGCAACGACGATGCTGTGTGTGTCACCTCAAGACTG 1288  
 QY 181 CysThrAsnAsnGlyLeuMetIleAlaTrpAsnGlyIleGluArgLeuArgAlaGlyLeu 200  
 Db 1289 TGACCTACCAATGACATCATATATGCA----- 1315  
 QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrgIuProLysCysProLeuGlyValAsp 220  
 Db 1316 -----TGATGCTCTTGGATGAC 1336  
 QY 221 ILeSerTyrgIuValGlyGluAlaSerIleLysValProGlnLeuLysMetGluIle 239  
 Db 1337 ATATCCAGAGAGATTGCGAGAGCTGCATAAAGTACCGCATTAATAATGGCACTT 1393  
 RESULT 7  
 LOCUS BE740611 701 bp mRNA linear EST 15-SEP-2000  
 DEFINITION 601595739P1 NIH\_MGC\_9 Homo sapiens CDNA clone IMAGE:3949640 5',  
 mRNA sequence.  
 BE740611  
 ACCESSION BE740611.1 GI:10154603  
 VERSION EST.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 701)  
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMNL at: Image.liml.gov  
 Plate: LHC614 row: n column: 09  
 High quality sequence stop: 701.  
 Location/Qualifiers  
 1..701  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3949640"  
 /tissue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_9"  
 /note="Organ: ovary; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-CDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.31e-95 Length: 701  
 Score: 908.00 Matches: 178  
 Percent Similarity: 58.91% Conservative: 3  
 Best Local Similarity: 97.21% Mismatches: 0  
 Query Match: 73.23% Indels: 2  
 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x BE740611 (1-701)  
 QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
 Db 155 CTGTTGGCAATGATGTTCAAGAGATTTCAGATTTCTCTTCTTGGAAATCTTTGGACATA 214  
 QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
 Db 215 GCAACGAGGACATGCTTACACAGTGGCAAGAGACTTTCTTATTAACATCCAGAG 274  
 QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnAspPhe 60  
 Db 275 TGTCTCACCATTAGATGTGTGGAAAGCCATAGACATTGGCCAAACAGAAATATGATTT 334  
 QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheTrgIy 80  
 Db 335 CATTTTGACATCAAACTCTTCATCATGCTTAAAAATTTGATTTTCTTTACTGGA 394  
 QY 81 LeuGlnHisValThrAspLysIleIleMetLys--LysGluLysGluGluGlyIleGlu 100  
 Db 395 CTTCAACACGTTACTGATTAATTAATTAATGAACAGAAACAGAAAGATTTAGGA 454  
 QY 100 YsgIyGlnIleLeuSerSerAlaAlaAspIleAlaIleThrValGlnHisThrMetAlaC 120  
 Db 455 AGGGGGAATCTCTGTTTTCAGCAGCAGACATTCCTCCACAGTACACACAAATGGCAT 514  
 QY 120 YhisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuAsp 140  
 Db 515 GTCATCTTGTTGAAGAAACACATCGGGCTATTCTGTTTGTAAAGACAGAGACTTGTAC 574  
 QY 140 TCGIAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPhenylleArg 160  
 Db 575 CTCAAATTAATGACAGTACTGTTGATCTGTGTGTGCCAATTAATCTTAATATCCGA 634  
 QY 160 rGAlaLeuGluLeuLeuThraAsnAlaThrGlnCysThrLeuLeuCysProProAlaArg 180  
 Db 635 GAGCTCTGGAATTTTAACAAACGACACAGTGCCTTTGTTGTCTCTCTCCAGAC 694  
 QY 180 euCys 181  
 Db 695 TATGC 699  
 RESULT 8  
 CF362328/c 658 bp mRNA linear EST 25-AUG-2003  
 LOCUS CF362328  
 DEFINITION 829596 MARC 3P1G Sus scrofa cDNA 3', mRNA sequence.  
 ACCESSION CF362328  
 VERSION CF362328.1 GI:34161882  
 KEYWORDS EST.  
 SOURCE Sus scrofa (pig)  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 REFERENCE 1 (bases 1 to 658)  
 AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,  
 Wise,T.P.L., Noneman,D.J., Wray,J.B. and Keele,J.W.  
 TITLE A second set of porcine ESTs from a pooled-tissue normalized  
 library  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Smith TP  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called with phred v0.020425.c and  
 trimmed with the aid of the trim\_alt option. Vector identified with  
 cross\_match v0.990329.  
 Plate: SRG8015 row: H column: 12  
 Seq primer: TAGAAGCAGCACTGCAGC.  
 Location/Qualifiers  
 1..658

```

/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3P1G"
/notice="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI"
library_made_with="RNA pooled from multiple tissues including brain, liver, muscle, placenta/endometrium, ovary, testes, and bone marrow."

```

**ORIGIN**

Alignment Scores:	
Pred. No.:	3, 07e-91
Score:	876.00
Percent Similarity:	93.68%
Best Local Similarity:	86.84%
Query Match:	70.65%
DB:	7
length:	658
Matches:	165
Conservative:	13
Mismatches:	12
Indels:	0
Gaps:	0

US-10-649-273-2\_COPY\_176\_414 (1-239) X CF362328 (1-658)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuLeuGlyValSerLeuAspIle 20  
Db 630 CTTTGGCATTAGTTAGAGAGAGTTTCAGATTTTCCTCTCTTGACAGCTTTTGACATA 5711  
QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
Db 570 GCACCAAGGTGACATGCTTGACAAAGGTAGCAGAAGCTTTCTTATTAACAATCCAGAG 5111  
QY 41 CysSerThrMetSerGlyGlyLysAlaIleGluHisLeuAlaLysGlnGlyAsnArgPhe 60  
Db 510 TGCCTCCACCAAGAGGTGGAGAGGCCATAGAACATTGTGTCCAAACAGGGAATTAAGTTG 4511  
QY 61 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerPheThrGly 80  
Db 450 CATTTGAATTTCCAACCTCCCAATGCACCTGATTAATAAATGTGATTTTCTTTTCGGA 3911  
QY 81 LeuGlnHisValThrAspLysIleIleMetLysGlyLysGlnGlyIleGluLys 100  
Db 390 CTTCAACAATGTTATTTGATTAGACAAATATNGCAGAAGCAAGAGAGAGATATTAGGAAG 3311  
QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120  
Db 330 GGGCAAAATCCCTGCTCTTCAGCTGCACAACTGCTGCGCAGTACAGCACACAGTAGCGCTGC 2711  
QY 121 HisLeuValLysArgGThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuPro 140  
Db 270 CATATTTGCAAAAGAAGCACTCATGTGCTATTCGTTTGGCAACAGAGAGCTTAATATGT 2111  
QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyGlyValAlaSerAsnPheTyrIleArgArg 160  
Db 210 CAAGTAAATGCAGTATTTGGTTCTATCTGGAGAGGTTCGCAAGTAACTTAATATACAAAA 1511  
QY 161 AlaLeuGlnIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180  
Db 150 GCCTTAGAAGTTTGACAAATGCACAACATGCACTGTGTGTGCTCTCCCTCCAGACTA 91  
QY 181 CysThrAspAsnGlyIleMetIleAlaTrp 190  
Db 90 TGCACGTATTAATGCAATTATGATTCATGG 61

RESULT 9  
CK941819/c

LOCUS	CK941819	637 bp	mRNA	linear	EST 15-MAR-2004
DEFINITION	06565407 BARC 10BOV Bos taurus CDNA clone 10BOV12_F24 3', mRNA sequence.				
ACCESSION	CK941819				
VERSION	CK941819.1	GI:45456199			
KEYWORDS	EST.				
SOURCE	Bos taurus (cow)				
ORGANISM	Bos taurus				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;				

REFERENCE	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
AUTHORS	Bovine; Bos.
TITLE	1 (bases 1 to 637) Sonstegard T.S., Van Tassell, C.P., Matukumalli, L.K., Harhay, G.P., Bosak, S., Rubenfield, M. and Gasbarre, L.C. Production of EST from cDNA libraries derived from immunologically activated bovine gut
JOURNAL	Unpublished. (2004)
COMMENT	Contact: Tad S. Sonstegard

**COMMENT**

**COMMENT**

Bovine Functional Genomics Laboratory  
Animal and Natural Resources Institute  
Bldg. 200 Ra2A BARC-East, Beltsville, MD 20705, USA  
Tel.: 3015048416  
Fax: 3015048414  
Email: ted@anri.barc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred  
0.000925 using options -trim at " -trim fasta. Vector identified  
by cross match using options -mismatch 12 -mismatch 18  
Plate: 12 row: F column: 24  
Seq primer: AGCGGATTAACATTTGACACGAG  
High quality sequence stop: 637.

**FEATURES**  
**Source**

**Source**

```

/organism="Bos taurus"
/mol_type="mRNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="10BOV12_F24"
/sex="Male"
/tissue_type="Pooled"
/dev_stage="Multiple"
/lab_host="DH10B T1 phage resistant"
/clone_lib="PARC 10BOV"
/notes="Organ: Small Intestine; Vector: pagen-1; Site:1;
ECORV; Site:2: NotI; Equimolar amounts of mRNA extracted
from proximal jejunums of 18 and 21 wk old steers, and
distal ileum of 14 day old calves. proximal jejunum
exposed to C. oncophora for 3 and 6 weeks, and distal
ileum exposed to C. parvum for 7 days"

```

US-10-649-273-2\_COPY\_176\_414 (1-239) x CK941819 (1-637)

Oy	51	AlaIysGlnIInclYAmnRgPheHIsplEapRlIeIyPProCHeuHIsHsAlaIyAsn	73
Db	632	GCCAAACAGGAAATGATTTGCACTTTTGATTTCCAGCTCCCAACGTCGTAAATAAT	57
Oy	74	CysAspHeSerPheThrGlyLeuGlnHisValThrAspLysIleIleMetLysIysGln	93
Db	572	TGTGATTTTCTTTTCTGCACTTCAACAGTTATGATGATTAATGCAAAAGAA	51
Oy	94	LysGIuGIuGIyIleGIuLysGIyGlnIleLeuSerSerAlaIAspIleAlaIthr	111
Db	512	AAAGAGGAAGGTTCGAGCAGGGCGAGTCTCTCTTCACGTCGCAACATTCGTTGGCG	45
Oy	114	ValGIuHisThrIleAlaCysHisIleuValLysArgThrHisArgAlaIleuPheCys	133
Db	452	GTGCAGCACCCGGGCGCTCCACATTCGCAAAAGAACACATCGTCTCTTCGTCTTCG	39
Oy	134	LysGlnIArgAspLeuLeuProGlnIAsnAsmAlaValIleuValAlaSerGIyIValAla	15
Db	392	AAGCAGAGAGGCTTCTTACGTCAGAGTAACGCACTGCTGTATCTGGAGGCGTCGCA	33
Oy	154	SerAsnPheIrrIleArgArgAlaLeuGluIleLeuThrAsmAlaThrGlnCysThrIleu	17



Db 332 AGTACTTATATATCCGAAAGCCCTGGAAATTGTGACCAATGACACAGTGCCTTTG 273  
 QY 174 LeuCySPProProCysLeuCyThraSPasnGlyIleMetIleAlaTrpAsnGlyIle 193  
 Db 272 CTGTGGCCGCCCCCAGACTTGTGACTGACACAGCGCTTATGATTCATGAAATGCTTT 213  
 QY 194 GlnArgLeuArgAlaGlyLeuGlyIleLeuHisAspIleGlyIleArgTrpGluPro 213  
 Db 212 GAAAGACTAGCTGCTGGCTGGGCTTTTACACACAGAAAGCATCCGCTACGACCA 153  
 QY 214 LysCySPProLeuGlyValAspIleSerIleGlyValAlaSerIleLysValPro 233  
 Db 152 AAAGTCTCTCTTGAGTATATCAAAAGAGTTGAGAGCTGCTATAAAGTGCAC 93  
 QY 234 GlnLeuLysMetGluIle 239  
 Db 92 AGATTAAATGAAGATT 75  
 RESULT 10  
 LOCUS BO961028 922 bp mRNA linear EST 21-AUG-2002  
 DEFINITION AGENCOURT\_8863711 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:6423902  
 5', mRNA sequence.  
 ACCESSION BO961028  
 VERSION BO961028.1 GI:22376506  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 922)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: rs9@bbs-rcmail.nih.gov  
 Tissue Procurement: DCTD/DRP/Gazdar  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: LICM2603 row: d column: 15  
 High quality sequence stop: 584.  
 Location/Qualifiers  
 1..922  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6423902"  
 /issue\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 18"  
 /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCAAGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC Library."

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyLysSerLeuAspIle 20  
 Db 291 CTGTGGCATTGATTCACAGAGATTTCAGATTTTCTGCTCTTGGAAGCTTTGGACATA 350  
 QY 21 AlaProGlyAspMetLeuAspLysValAlaArgArgLeuSerLeuIleLysHisProGlu 40  
 Db 351 GCACCAAGGAGATGCTTGAACAAGTGGCAAGAAGACTTTCTTATTAACAATCCAGAG 410  
 QY 41 Cys-SerThrMetSerGlyGlyLysAlaIleGlnHisLeuAlaLysGlnGlyAsnArgph 60  
 Db 411 TGCNTCCACCAAGAGTGGTGAAGAACCATGAAACATTTGGCCAAACAGAAATGAGTT 470  
 QY 60 HisPheAspIleLysProProLeuHisHisAlaLysAsnCysAspPheSerThrGln 80  
 Db 471 TCATTTTGAATCAAAACCTCCCTTGATCATCTAAATAATTTGATTTTCTTTACTGG 530  
 QY 80 LysGlnHisValThrAspLysIleIleMetLysGlyLysGlyGlyIleGlyIleGlyIle 100  
 Db 531 ACTTCACACAGTTACTGATTAATAATATGAAGAAAAAGAGAAAGATTTGAGAA 590  
 QY 100 GlnGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetLac 120  
 Db 591 AGGCAAAATCCTGTCTTCAGACAGACATTCCTGCACAGTACACACACATGCGCATG 650  
 QY 120 HisLeuValLysArgThrHisArgAlaIleLeuPheCysLysGlnArgAspLeuLeuPr 140  
 Db 651 TCATCTTGTGAAGAAAGAA-CATCGGCTATTTCTGTTTGTACACAGAGACTTGGTACC 709  
 QY 140 GlnAsnAsnAlaValLeuValAlaSer-GlyGlyValAlaSerAsnPheTrpIleArgA 160  
 Db 710 TCAAAATATATGACAGTACTGGTGCATCTTGGGGGGTGGCAAGTACTTATATCCCG 769  
 QY 160 ArgAlaLeuGluIleLeuThr-AsnAlaThrGlnCysThr-LeuLeuCySPProProArg 179  
 Db 770 AAACCTCGGAAAAAATTACAAACCCACACGCGGACCTTTGTTGGTCCCTCCCAA 829  
 QY 179 GlnCySPThr-AspAsnGlyIleMetIleAlaTrp-----AsnGlyIleGluArg---L 196  
 Db 830 ACTATGCACTGATTAATAGG---CATTAATGATGCTGGGGAAGGAAATTTGAATAATA 886  
 QY 196 euArgAlaGlyLeuGly 201  
 Db 887 CCATGCTGGCTTGGGG 903  
 RESULT 11  
 LOCUS AV602901/c 545 bp mRNA linear EST 27-NOV-2001  
 DEFINITION AV602901 Bos taurus kidney fetus Bos taurus cDNA clone EIKI013A07  
 3', mRNA sequence.  
 ACCESSION AV602901  
 VERSION AV602901.1 GI:9725227  
 KEYWORDS EST.  
 SOURCE Bos taurus (cow)  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.  
 1 (bases 1 to 545)  
 Establishment of a high throughput EST sequencing system using  
 poly(A) tail-removed cDNA libraries and determination of 36,000  
 bovine ESTs  
 Nucleic Acids Res. 29 (22), E108 (2001)  
 JOURNAL MEDLINE  
 PUBMED 11713328  
 COMMENT Contact: Yoshikazu Sugimoto  
 Animal Genetics Division  
 Shitakawa Institute of Animal Genetics  
 Odakura, Nishigo, Nishi-shitakawa, Fukushima 961-8061, Japan  
 Tel: 81-248-25-5641  
 Fax: 81-248-25-5725

Email: kazunugi@cocoa.ocn.ne.jp  
Single pass sequencing.  
This clone was obtained from a polyA-deleted cDNA library.

## FEATURES

Source  
Location/Qualifiers

1..545  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/clone="BK1013A07"  
/tissue\_type="kidney"  
/dev\_stage="fetus"  
/lab\_host="DH10B"  
/clone\_lib="Bos taurus kidney fetus"  
/note="Vector: pZ1; Site 1: SalI; Site 2: NotI; Poly A  
was deleted from a NotI site"

## ORIGIN

## Alignment Scores:

Pred. No.: 3,18e-81 Length: 545  
Score: 789.00 Matches: 147  
Percent Similarity: 90.11% Conservative: 17  
Best Local Similarity: 80.77% Mismatches: 17  
Query Match: 63.63% Indels: 1  
Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x AV602901 (1-545)

QY 57 G1YASNRGPHNHPHAPRLELYSPROFLEUHNHSHLS-ALALYASNCYASAPH 76  
DB 544 GGAATATGATTTGATTTGATTTCCANCNTCCAGCAACGGTCTGTAATAATTTGATTT 485  
QY 76 ESEPHATRG1YLEUG1NHISVALTHRASPLATLEIEMETLYSLYSG1ULYSG1 96  
DB 484 TTTCTTTCTGCACTTCAACAGTTATTGATGATGATGATGCAAAAGAAANAGGA 425  
QY 96 UG1Y1EG1ULYSG1ULYSG1LEU5ERSERX1AALASPILEALATHRVALG1NH1 116  
DB 424 AGCATGAGAGAGGGGCGAGTCTCTTCAGCTCGGACATTCGTTCGGTCCAGCA 365  
QY 116 ETHMETALACYSHISLEUVALYASXTHRTHISARG1AILEUAPHECYALYSG1NAR 136  
DB 364 CACCGTCCCTGCACATTCGAAAGAAACATGCTCTCTGTTCTGCAACAGAG 305  
QY 136 GASP1LEU5ERPROG1NAPENAP1AVALLEUVAL1ASERG1YVAL1ASERAPH 156  
DB 304 AGGCTTCTTACATGACGATTAACGAGTACTGTTATCTGAGGCGTGCAGATTA 245  
QY 156 ETY11EARGARG1ALEUG1ULYLEU5THRASN1ATHRG1NYSRTHLEUCYAPR 176  
DB 244 ATATATGCGAAGAACCCCTGGAATTTGACCAATGCAACAGTGCACATTCGTGCCC 185  
QY 176 OPTOPROARGLEUCYSTRASAPENG1Y1LEMET1EALATTPAENG1Y1LEGLUARG1E 196  
DB 184 GCCCCCAAGACTCTGCACTGACCAACGGCGTATGATTCATGATGATGGTGAAGACT 125  
QY 196 UARG1AG1YLEUG1Y1LEUHNHISAPR1EG1UG1Y1LEAUG1Y1RGIUPROLYCYAPR 216  
DB 124 AGGCTGCGCTTGGCATTTTACACACAGAAAGCAATCCGCTACGAACCAAAATGTC 65  
QY 216 OLEUG1VALASPILESERLYSG1ULYVAL1G1YUL1ASER1Y1EVAL1PROG1LEU 236  
DB 64 TCTTGAGATGATATATCAAAAGAGTTGAGAGAGCTGCTATTAAGTCCCAAGATTAA 5  
QY 236 Smet 237  
DB 4 AATG 1

RESULT 12  
LOCUS CB852881 490 bp mRNA linear EST 22-APR-2003  
DEFINITION UI-CF-FNO-afo-b-03-0-UI-s1 UI-CF-FNO Homo sapiens cDNA clone  
ACCESSION CB852881 UI-CF-FNO-afo-b-03-0-UI 3', mRNA sequence.

VERSION CB852881.1 GI:30047942  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Fukuyama, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 490)

JOURNAL  
MEDLINE  
PUBMED  
Genome Res. 6 (9), 791-806 (1996)

## COMMENT

97044477  
8889548  
Contact: McCray, PB  
McCrays Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welch, University of Iowa  
CDNA library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.research.com) or from Open Biosystems  
(www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA-No.

## FEATURES

Source

Location/Qualifiers

1..490  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-FNO-afo-b-03-0-UI"  
/tissue\_type="Human lung Epithelial cells"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-FNO"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-CF-FNO is a subtracted cDNA library derived from two  
normalized Human lung epithelial cell libraries (EN1 and  
DUI). The library was subtracted according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. For additional information, contact:  
bento-soares@uiowa.edu  
TAG\_SEQ=None found"

## ORIGIN

## Alignment Scores:

Pred. No.: 1.05e-80 Length: 490  
Score: 784.00 Matches: 148  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 63.23% Indels: 0  
Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x CB852881 (1-490)

QY 69 H1SH1SALAYASNCYASAPRPHESERPHETRG1YLEUG1NHISVALTHRASPLYS1LE 88  
DB 40 CATCATGCTGTAATAATTTGATTTCTTTACTGCACTTCAACAGCTTACGTATTAATA 99  
QY 89 I1EMETLYSLYSG1ULYSG1UG1Y1LEGLUARG1Y1LEU5ERSER1AAL1A 108  
DB 100 ATATGTAAGAAAGGAGGAGGATATGAGAAAGGCAAAATCTGCTTCAAGAGCA 159  
QY 109 AAPI1EALATHRVALG1NH1SRTHMETALACYSHISLEUVALYASATGTH1SARG 128  
DB 160 GACATTGCTGCCACAGTACAGCACAAATGGCAATGTCATCTGTGAAGAAACACATCGG 219  
QY 129 A1A1LEU5PHCYSLYSG1NARGAPR1EULEU5PROG1NAPENAP1AVALLEUVAL1A 148

Db 220 GCATTTCTGTTTGTAAAGCAGAGACTTGTTACTCAAAATATAGCACTACTGTTGGC 279

QY 149 SerGIyGIyValAlaSerAspPheTyrIleArgAlaLeuGluIleuThrAspAla 168

Db 280 TCTGGTGTGTGCGAGATTAATCTTATCCAGAGCTCTGGAATTTTAAACAAAGCA 339

QY 169 ThrGlnCysThrLeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIle 188

Db 340 ACACAGTGCATTTGTGTGTCTCTCCAGACATATGACATGATATGATATGATT 399

QY 189 AlaTTPanGlyIleGluArgLeuArgAlaGlyIleuGlyIleuHisAspIleGlyIle 208

Db 400 GCATGGAATGGATTTGAAGAAGCTAGCTGCTGGCATTTTACATGACATAGAGGC 459

QY 209 ILeaArgTyrGluProLeuCysPro 216

Db 460 ATCCGCTATGAAACCAAAATGTCCTC 483

RESULT 13

LOCUS CF257246 822 bp mRNA linear EST 07-AUG-2003

DEFINITION phao08\_g02 PHA-activated splenocytes Gallus gallus cdna, mRNA

ACCESSION CF257246

VERSION CF257246.1 GI:33490501

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 822)

AUTHORS Wiltzell, H., Bed'Hom, B., Morin, V., Young, J.R., Whitaker, C.J., Chausse, A.M. and Zoorob, R.

TITLE A collection of chicken ESTs from activated immune cells

JOURNAL Unpublished (2003)

COMMENT Contact: Zoorob R

CNRS UPR 1983

7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France

Tel: 33 1 49 58 35 00

Fax: 33 1 49 58 33 81

Email: zoorob@vjf.cnrs.fr.

FEATURES

source location/Qualifiers

1..822

/organism="Gallus gallus"

/mol\_type="mRNA"

/db\_xref="taxon:9031"

/cell\_type="Splenocytes"

/clone\_lib="PHA-activated splenocytes"

/note="Vector: pTIP1EX2"

ORIGIN

Alignment Scores:

Pred. No.: 2.85e-80 Length: 822

Score: 783.00 Matches: 150

Percent Similarity: 79.57% Conservative: 37

Best Local Similarity: 63.83% Mismatches: 48

Query Match: 63.15% Indels: 1

DB: 7 Gaps: 0

US-10-649-273-2\_COPY\_176\_414 (1-239) x CF257246 (1-822)

QY 1 LeuLeuAlaLeuValGlnGlyValSerAspPheLeuLeuGlyIleSerLeuAspIle 20

Db 119 ATCTTGGAGTAGACAGAGAGTTTCATTTCTTCTTGCTTGAGACATGATATCA 178

QY 21 AlaProGlyAspMetLeuAspIleValAlaArgAlaGluSerLeuIleIleHisProGlu 40

Db 179 GCACCAAGGTGACATGTTTCACTTACCTACCAAGAGCTCTTTAGTGAAGCACCAGGAG 238

QY 41 CysSerThrMetSerGlyIleValAlaIleGluHisLeuAlaIleGlnGlyIleAsnArgPhe 60

Db 239 TGCACAGCATGCGCGGGGAGGACATAGACACCTGCTCAACCGAGACTGGCAA 298

QY 61 HisPheAspIleLeuSerProLeuHisPheAlaIleAsnGlyIleSerPheSerThrGly 80

Db 299 CAGTACCTTTCAGACTTCCATGCAACAGATATGATATGATATTTCTTCTCCGGA 358

QY 81 LeuGlnHisValThrAspIleIleIleMetIleGlyIleGluGluGluGluGlu 100

Db 359 CTTCCAGAGCTTGTCAACAAAGCATTTCTCAGAAAGAAAAGAAAGATATTCAGAA 418

QY 101 GlyGlnIleLeuSerSerAlaAlaAspIleAlaAlaThrValGlnHisThrMetAlaCys 120

Db 419 GGGGAATCTCTGCTCGCTTAAAGACATCGCTGCTGCACAGCAGTATGCTGCT 478

QY 121 HisLeuValIleSerThrHisArgAlaIleLeuPheCysIleGlnArgAspLeuPro 140

Db 479 CATATTATCCAGCGGACACCGAGCATGCTTCTGCAATGAAAACAGCATATTATTA 538

QY 141 GlnAsnAsnAlaValLeuValAlaSerGlyIleValAlaSerAspPheTyrIleArgArg 160

Db 539 CCAAAAGCTGCACCTGCTGTTATCAGAGAGCTTGCAGATATCATCAGAAA 598

QY 161 AlaLeuGluIleLeuThrAsnAlaThrGlnCysThrLeuLeuCysProProArgLeu 180

Db 599 GGAATCAGACTCTGCGCAATGCAACGGTTTGTCTTCTGCTCTCTCCAAAGCTG 658

QY 181 CysThrAspAsnGlyIleMetIleAlaIleAsnGlyIleGluArgLeuArgAlaGlyLeu 200

Db 659 TGCACCATTAATGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 718

QY 201 GlyIleLeuHisAspIleGluGlyIleArgTyrGluProLeuCysProLeuGlyValAsp 220

Db 719 GGAATTTATACAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 778

QY 221 IleSerIleGluValGlyIleValAlaSerIleIleValProGluLeu 235

Db 779 ATTTCCAAAGAGTTGAA-GAGGATTCATCAAAAGTCCAAAGACTA 822

RESULT 14

LOCUS BUI27463 866 bp mRNA linear EST 25-NOV-2002

DEFINITION 603114407F1 CSBQHL20 Gallus gallus cdna clone CHEST6m2 5', mRNA

ACCESSION BUI27463

VERSION BUI27463.1 GI:25338728

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 866)

AUTHORS Boardman, P.B., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Ford, M.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

TITLE A Comprehensive Collection of Chicken cDNAs

JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)

MEDLINE 22335534

PUBMED 12445392

COMMENT Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology

(UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source location/Qualifiers

1..866

/organism="Gallus gallus"

/mol\_type="mRNA"

/strain="White Leghorn, Hisex"

/db\_xref="taxon:9031"

/clone="CHST6m2"



```
Db 121 GCGGTACAGCAGCGAACGCGCTGCCACCTTGCGAAAGAACACATCGTCTATTCGTTT 180
QY 133 CyelysElnArgAspLeuLeuProGlnAsnAsnAlaValLeuValAlaSerGlyGlyVal 152
Db 181 TGGCAGCAGAAAAATTGCTATCTCCAGCTAACGCAATATTAAGTGTGTGAGAGTGT 240
QY 153 AlSerAspPheTyrIleArgArgAlaLeuGluIleLeuThrAsnAlaThrGlnCysThr 172
Db 241 GCAGTAACTTGTCATCCGAGAGCATTTGAAATTGTAGCAAAATGCACACATGCACT 300
QY 173 LeuLeuCysProProProArgLeuCysThrAspAsnGlyIleMetIleAlaTrpAsnGly 192
Db 301 TTGTTGTGTCCCCCTCCGAGACTGTGCACTGACATGGATCATGATTCATGCAATGGA 360
QY 193 IleGluArgLeuAlaGlyIleLeuHisAspIleGluGlyIleArgTyrGlu 212
Db 361 ATTGAAAGATTACGTGCTGGCTTGGCATTTTACATGATGTAGAGAGACATCCGATACGA 420
QY 213 ProLysCysProLeuGlyValAspIleSerLysGluValGlyGluAlaSerIleLysVal 232
Db 421 CCAAAATGTCTCTCGAATATGACATATCCAGAGAGTTGCGAAGCTGCCATAAAGTA 480
QY 233 ProGlnLeuLysMetGluIle 239
Db 481 CCAAGATTAAAAATGACACTT 501
```

Search completed: February 16, 2005, 21:04:55  
Job time : 2847.68 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 10:46:17 ; Search time 40.5261 Seconds  
(without alignments)  
3019.954 Million cell updates/sec

Title: US-10-649-273-2\_COPY\_176\_414  
Perfect score: 1240  
Sequence: 1 LIALVQGVSDPFLIGKSLDI.....DISKVGSAIKVPOLKMEI 239

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Uniprot\_03: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1240	100.0	414	2	096EV9 homo sapien
2	1213.5	97.9	439	2	09H4B0
3	1086	87.6	414	2	06PEB4 mus musculu
4	1079	87.0	414	2	06BLB6 mus musculu
5	1071	86.4	414	2	09D0N0
6	983	79.3	467	2	06AYN7 rattus norv
7	960	77.4	364	2	096NH5 homo sapien
8	751	60.6	404	2	08JFW3 brachydanio
9	747	60.2	404	2	08JF7 brachydanio
10	361.5	29.2	401	2	07Q9I8 anopheles g
11	332.5	26.8	409	2	0960S6 dirosophila
12	332.5	26.8	409	2	09VMD6 dirosophila
13	324	26.1	480	2	022145 arabidopsis
14	299.5	24.2	255	2	0677H2 hyacinthus
15	262.5	21.2	360	2	092LH8 rhizobium m
16	257	20.7	346	2	07VXN4 bordetella
17	254	20.5	335	2	073H71 wolbachia p
18	253	20.4	362	2	098BI6 acinetobact
19	252.5	20.4	340	2	06FCR9 acinetobact
20	252	20.3	346	2	07W6E8 bordetella
21	251	20.2	346	2	07W134 rhodospseudo
22	246.5	19.9	389	2	068XR3 rickettsia
23	245	19.8	387	2	08UC47 agrobacteri
24	244	19.7	365	2	07CWJ8 agrobacteri
25	244	19.7	366	2	08YJBI bruceella me
26	241.5	19.5	359	2	08FY15 bruceella su
27	241.5	19.4	342	2	09LJ11 pasteurella
28	239	19.3	251	2	093FJ2 cowdria rum
29	238	19.2	387	1	09E88 rickettsia
30	237.5	19.2	367	2	09ABZ9 caulobacter

32	236.5	19.1	357	2	089WM1 bradyrhizob
33	234.5	18.9	344	2	07PAG7 rickettsia
34	233.5	18.8	339	2	06LVI10 photobacter
35	231	18.6	341	2	07NIB3 chromobacte
36	231	18.6	341	2	091SV7 pseudomonas
37	231	18.6	364	2	06GIR3 bartonella
38	230.5	18.6	353	2	092UK6 candidatus
39	228.5	18.4	343	2	07VQO9 candidatus
40	224	18.1	337	2	082XN2 nitrosomona
41	224	18.1	342	1	GCP HAEIN haemophilus
42	223.5	18.0	344	2	065EP0 mammatoma
43	221.5	17.9	323	2	0291S3 archaeglob
44	219.5	17.7	347	2	08XX97 ralsomona s
45	217	17.5	341	2	08ES16 oceanobacil

## ALIGNMENTS

RESULT 1  
ID 096EV9 PRELIMINARY; PRT; 414 AA.  
AC 096EV9;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE OSCGRL1 protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schler G.D.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schler G.D.,  
Altschul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,  
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Mallory S.J.,  
Raha S.S., Loguercio N.A., Peters G.J., Abramson R.D., Mullan S.J.,  
Boak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
Krzyszinski W.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,  
Jones S.J., Marra M.A.,  
"Generation and initial analysis of more than 15,000 full-length human  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RX Strausberg R.;  
RA Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
EMBL: BC011904; AAH11904.1; --  
DR MEROPS; M22.004; --  
DR GO; GO:0008450; F:O-ialloglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F:Zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009050; Peptidase M22.  
DR InterPro; IPR009180; Peptidase M22\_Ostia1.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PIRSF; PIRSF004537; Ostia1g1c; 1.  
DR PRINTS; PR00789; OSIALOPTASR.  
DR Prodom; P0002367; Peptidase M22; 1.  
DR TIGRFAMs; TIGR00329; gcp\_1.  
SQ SEQUENCE 414 AA; 45122 MW; AS36B333F5C6B8DD CRC64;

Query Match 100.0%; Score 1240; DB 2; Length 414;  
Best Local Similarity 100.0%; Pred. No. 5,6e-101;  
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIALVQVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRF 60  
DB 176 LIALVQVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRF 235  
QY 61 HFDIKPPLHAKNCDPFTGLQHVTDKILMKKEKEGEGILSSADIAATVQHTMAC 120  
DB 236 HFDIKPPLHAKNCDPFTGLQHVTDKILMKKEKEGEGILSSADIAATVQHTMAC 295  
QY 121 HLKRTTRAILFCQKORDLLPNNNAVLVAGSVAFYIRRALEILTNAQCTLLCPPL 180  
DB 296 HLKRTTRAILFCQKORDLLPNNNAVLVAGSVAFYIRRALEILTNAQCTLLCPPL 355  
QY 181 CTDNGIMIAMNGIERLRAGLILHDIGIRYEPKCPGLVDISKVGBASIKVPOLKMEI 239  
DB 356 CTDNGIMIAMNGIERLRAGLILHDIGIRYEPKCPGLVDISKVGBASIKVPOLKMEI 414

## RESULT 2

Q9H4B0 PRELIMINARY; PRT; 439 AA.  
AC 09H4B0  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Putative sialoglycoprotease type 2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Chen J.M., Fortunato M., Barrett A.J.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
EMBL; AJ295148; CACI466.1; -.  
DR MEROPS; M22.004; -.  
DR Genew; HGNC:23075; OSGEPL1.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009180; Peptidase M22.  
DR Pfam; PF00814; Peptidase M22; 1.  
DR PRINTS; PIRSF004537; OsiAlgic\_ptcde; 1.  
DR PRODOM; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
KW Protease.  
SQ SEQUENCE 439 AA; 48040 MW; 44849372C784E41F CRC64;

Query Match 97.9%; Score 1213.5; DB 2; Length 439;  
Best Local Similarity 90.2%; Pred. No. 1.3e-98;  
Matches 238; Conservative 0; Mismatches 1; Indels 25; Gaps 1;

QY 1 LIALVQVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRF 60  
DB 176 LIALVQVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRF 235  
QY 61 HFDIKPPLHAKNCDPFTGLQHVTDKILMKKEKEGEGILSSADIAATVQHTMAC 98  
DB 236 HFDIKPPLHAKNCDPFTGLQHVTDKILMKKEKEGEGILSSADIAATVQHTMAC 295  
QY 99 ---EKQGLSSADIAATVQHTMACHLVYRTRAILFCQKORDLLPNNNAVLVAGSVASN 155  
DB 296 CRVEKGLILSSADIAATVQHTMACHLVYRTRAILFCQKORDLLPNNNAVLVAGSVASN 355  
QY 156 FYIRRALEILTNAQCTLLCPPLCTDNGIMIAMNGIERLRAGLILHDIGIRYEPKC 215

DB 356 FYIRRALEILTNAQCTLLCPPLCTDNGIMIAMNGIERLRAGLILHDIGIRYEPKC 415  
QY 216 PLGVDISKVGBASIKVPOLKMEI 239  
DB 416 PLGVDISKVGBASIKVPOLKMEI 439

## RESULT 3

Q9EBE4 PRELIMINARY; PRT; 414 AA.  
ID 09EBE4  
AC 09EBE4  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CZECH II; TISSUE=Mammary tumor;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schler G.D.,  
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsien P.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stachenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Abramson R.D., Mullahy S.J.,  
RA Raha S.S., Louellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalins D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Maier M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CZECH II; TISSUE=Mammary tumor;  
RA Strausberg R.;  
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
EMBL; BC058172; AAH58172.1; -.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009180; Peptidase M22.  
DR InterPro; IPR009180; Peptidase M22.  
DR Pfam; PF00814; Peptidase\_M22; 1.  
DR PIRSF; PIRSF004537; OsiAlgic\_ptcde; 1.  
DR PRINTS; PIRSF004537; OsiAlgic\_ptcde; 1.  
DR PRODOM; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 414 AA; 44962 MW; 0418AD93C6FEBCAE CRC64;

Query Match 87.6%; Score 1086; DB 2; Length 414;  
Best Local Similarity 85.8%; Pred. No. 2.2e-87;  
Matches 205; Conservative 15; Mismatches 19; Indels 0; Gaps 0;

QY 1 LIALVQVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRF 60  
DB 176 LIALVQVSDPFLILGKSLDIAFGDMLDKVARRLSLIKHPECSTMSGKAIIEHLAKQGNRF 235  
QY 61 HFDIKPPLHAKNCDPFTGLQHVTDKILMKKEKEGEGILSSADIAATVQHTMAC 120

Db 236 HFTINPMPQNAKCDPSFTGLQHTDKLITNKEKEGIEKGQILSSADIAAHOHATAC 295  
 Qy 121 HLVRTRHRAILFCQKODLIPONNAVVASGVASNFYIRALEILTNATQCTLLCPPL 180  
 Db 296 HLAKTRHRAILFCQKONLSPANNAVVGASVSNLYIRKALEIVANATQCTLLCPPL 355  
 Qy 181 CTDNGIMIAMNGIERLRAGLIGLHDIEGIRYEPKPLGVDSISKEYGEASIKVPOLKMEI 239  
 Db 356 CTDNGIMIAMNGIERLRAGLIGLHDIEGIRYEPKPLGVDSISKEYGEASIKVPOLKMEI 414  
 RESULT 4  
 Q8BLB6 PRELIMINARY; PRT; 414 AA.  
 AC Q8BLB6;  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)  
 DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length  
 DE enriched library, clone: B230219017 product: similar to PUTATIVE  
 DE SIALOGLYCOPROTEASE TYPE 2.  
 GN Name-Osgp11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RT Meth. Enzymol. 303:19-44(1999).  
 [2]  
 SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RT Nature 409:685-690(2001).  
 [3]  
 SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RA The RIKEN Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RT Nature 420:563-573(2002).  
 [4]  
 SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RT Genome Res. 10:1617-1630(2000).  
 [5]  
 SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitanaka T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujisaki S., Inoue K., Togawa Y., Izawa M., Ohara B., Watanabe K.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RT Genome Res. 10:1757-1771(2000).  
 [6]  
 SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RA Adachi S., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirose T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Koyama S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akaiwa S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK045669; BAC32450.1; -  
 DR MEROBS; M22.004; -  
 DR MGD; MGI:1919335; Osgp11.  
 DR GO; GO:0008450; F-0-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F-peptidase activity; IEA.  
 DR GO; GO:0008270; Zinc ion binding; IEA.  
 DR GO; GO:0006508; P-proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009095; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc.pptds; 1.  
 DR PRINTS; PR00789; OSIALOPRASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFS; TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SEQUENCE 414 AA; 44927 MW; 9A2D27B576C7DB24 CRC64;  
 Query Match 87.0%; Score 1079; DB 2; Length 414;  
 Best Local Similarity 85.8%; Pred. No. 9,1e-87;  
 Matches 205; Conservative 14; Mismatches 20; Indels 0; Gaps 0;  
 Qy 1 LIALVGVSDPFLIGKSLDIPAGMDKVARRLSLIKHEPCSTMSGKALIEHAKQGNRP 60  
 Db 176 LIALVGVSDPFLIGKSLDIPAGMDKVARRLSLIKHEPCSTMSGKALIEHAKQGNRP 235  
 Qy 61 HEDIKRPLHAKKCDPSFTGLQHTDKLITNKEKEGIEKGQILSSADIAAHOHATAC 120  
 Db 236 HFTINPMPQNAKCDPSFTGLQHTDKLITNKEKEGIEKGQILSSADIAAHOHATAC 295  
 Qy 121 HLVRTRHRAILFCQKODLIPONNAVVASGVASNFYIRALEILTNATQCTLLCPPL 180  
 Db 296 HLAKTRHRAILFCQKONLSPANNAVVGASVSNLYIRKALEIVANATQCTLLCPPL 355  
 Qy 181 CTDNGIMIAMNGIERLRAGLIGLHDIEGIRYEPKPLGVDSISKEYGEASIKVPOLKMEI 239  
 Db 356 CTDNGIMIAMNGIERLRAGLIGLHDIEGIRYEPKPLGVDSISKEYGEASIKVPOLKMEI 414  
 RESULT 5  
 Q9DON0 PRELIMINARY; PRT; 414 AA.  
 ID Q9DON0;  
 AC Q9DON0;  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)  
 DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length  
 DE enriched library, clone:2610001M19 product: similar to PUTATIVE  
 DE SIALOGLYCOPROTEASE TYPE 2.  
 GN Name-Osgp11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo body;  
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RT Meth. Enzymol. 303:19-44(1999).



RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN PANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA the FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Komno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
 RT "Normalization and subtractions of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Komno H., Akiyama J., Nishi K., Kitsuana T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama S., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,  
 RA Okazaki Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multichannel sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
 RA Imocanti K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
 RA Kawai J., Kojima Y., Komno H., Kouda M., Koya S., Kuribara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Nunazaki R., Ono M.,  
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Miyamatsu M., Hayashizaki Y.,  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN EMBL: AK011265; BAB37506.1; -  
 DR MEROPS: M22.004; -  
 DR MGD; MGI:1919335; Oseep1.  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR00905; Peptidase\_M22.  
 DR InterPro: IPR009180; Pept\_M22\_Ost191.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Ost191c.ptds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR Prodom; PD002367; Peptidase\_M22; 1.  
 DR TIGRPMs; TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SSQENCE 414 AA; 44999 MM; 999BC689944DB24 CRC64;  
 Query Match 86.4%; Score 1071; DB 2; Length 414;  
 Best Local Similarity 85.4%; Pred. No. 4,6e-86;  
 Matches 204; Conservative 14; Mismatches 21; Indels 0; Gaps 0;

QY 1 LIALVGVSDPILIGKSLDIPAGDMLDKVARRSLIKHPECSTMSGKATIEHLAKOGRNF 60  
 Db 176 LIALVGVSDPILIGKSLDIPAGDMLDKVARRSLIKHPECSTMSGKATIEHLAKOGRNF 235  
 QY 61 HPDIKEPLHAAKNCPSFTGLQHTVDKIMKEKEGIEKGQILSSAADIAATVQHTMAC 120  
 Db 236 HFTINPPMQNANKNCPSFTGLQHTVDKIMKEKEGIEKGQILSSAADIAATVQHTMAC 295  
 QY 121 HLVKRTIRAILPCKQPDLLPNNNAVIVASGVASNYIRRAAILTNATQCTLLCPPRL 180  
 Db 296 HIAKRTIRAILPCKQPDLLPNNNAVIVASGVASNYIRRAAILTNATQCTLLCPPRL 355  
 QY 181 CTDNGIMTMNGIERPAGILHDIEGIRYEPKCPGVDSKEVGBASIKVPOLKMEI 239  
 Db 356 CTDNGIMTMNGIERPAGILHDIEGIRYEPKCPGVDSKEVGBASIKVPOLKMEI 414  
 RESULT 6  
 Q6A9N7 PRELIMINARY; PRT; 467 AA.  
 AC Q6A9N7; 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX PubMed=12477932; DOI=10.1073/pnas.2426038993;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Muljany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gamarane P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalins D.E., Schermer A., Schein J.E.,  
 RA Jones S.J., Maitra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Director MGC Project;  
 RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.  
 RN EMBL: BC078974; AAH76974.1; -  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR00905; Peptidase\_M22.  
 DR InterPro: IPR009180; Pept\_M22\_Ost191.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Ost191c.ptds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR Prodom; PD002367; Peptidase\_M22; 1.  
 DR TIGRPMs; TIGR00329; gcp; 1.  
 DR PROSITE; PS00435; PROKIDASE\_1; UNKNOWN\_1.  
 KW Hypothetical protein.

FT NON\_TER 467 467  
 SQ SEQUENCE 467 AA; 50799 MW; 474EB1B1959B8AC0 CRC64;  
 Query Match 79.3%; Score 983; DB 2; Length 467;  
 Best Local Similarity 87.9%; Pred. No. 3.1e-78;  
 Matches 186; Conservative 9; Mismatches 17; Indels 0; Gaps 0;  
 QY 1 LIALVGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKGNRF 60  
 DB 176 LIALVGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKGNRF 235  
 QY 61 HFDIKPPLHAAKNCDFSTGLQHTVDKIIIMKKEEGIEKGQILSSADIAATVOHTMAC 120  
 DB 236 HFTINPMQNAKNCDFSTGLQHTVDKIIIMKKEEGIEKGQILSSADIAAATVOHTMAC 295  
 QY 121 HLYKTRTHAILFCCKORDLIPONNAVIVASGVASNFYIRALIEILTNATCTLLCPPRL 180  
 DB 296 HLYKTRTHAILFCCKORDLIPONNAVIVASGVASNFYIRALIEILTNATCTLLCPPRL 355  
 QY 181 CTDNGIMIAMNGIERLRAGLGLHDIEGIRYEPK 214  
 DB 356 CTDNGIMIAMNGIERLRAGLGLHDIEGIRYEPK 389  
 RESULT 7  
 Q96NH5 PRELIMINARY; PRT; 364 AA.  
 AC Q96NH5;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein FLJ30879.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=Brain;  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,  
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
 RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa B., Omura Y.,  
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
 RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
 RA Tanai H., Kimata M., Watanabe M., Hirooka S., Chida Y., Ishida S.,  
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
 RA Kanehori K., Takahashi-Fujii A., Hara R., Takeuchi K., Arita M., Imose N.,  
 RA Nishizaki K., Yuki H., Oshima A., Sasaki N., Aotsuma S.,  
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki M., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirose M., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Ohtsuki R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita K.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 cDNAs.";  
 PT cDNA;  
 DR EMBL AK055441; BAB70923.1; -  
 DR GENBANK M22.004; -  
 DR MEROPS; M22.004; -

DR GO; GO:0008450; P:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000905; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22\_1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptds; 1.  
 DR PRINTS; PR00789; OSIALOPRASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRPFAM; TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SEQUENCE 364 AA; 39528 MW; E0E605A07D0EC3D6 CRC64;  
 Query Match 77.4%; Score 960; DB 2; Length 364;  
 Best Local Similarity 98.4%; Pred. No. 2.5e-76;  
 Matches 186; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 LIALVGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKGNRF 60  
 DB 176 LIALVGVSDPFLLGKSLDIAPGMDLKVARRSLIKHPECSTMSGKAIIEHLAKGNRF 235  
 QY 61 HFDIKPPLHAAKNCDFSTGLQHTVDKIIIMKKEEGIEKGQILSSADIAATVOHTMAC 120  
 DB 236 HFDIKPPLHAAKNCDFSTGLQHTVDKIIIMKKEEGIEKGQILSSADIAATVOHTMAC 295  
 QY 121 HLYKTRTHAILFCCKORDLIPONNAVIVASGVASNFYIRALIEILTNATCTLLCPPRL 180  
 DB 296 HLYKTRTHAILFCCKORDLIPONNAVIVASGVASNFYIRALIEILTNATCTLLCPPRL 355  
 QY 181 CTDNGIMIAM 189  
 DB 356 CTDNGIMIAM 364  
 RESULT 8  
 Q8JFW3 PRELIMINARY; PRT; 404 AA.  
 AC Q8JFW3;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE SI-42211013.4 (Novel glycoprotease).  
 GN Name=ds272B14.6;  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Babage A.;  
 RL Submitted (SBB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL591593; CAD43471.1; -  
 DR MEROPS; M22.004; -  
 DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000905; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22\_1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptds; 1.  
 DR PRINTS; PR00789; OSIALOPRASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRPFAM; TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SEQUENCE 404 AA; 43956 MW; 3A6B1DD050737E35 CRC64;  
 Query Match 60.6%; Score 751; DB 2; Length 404;  
 Best Local Similarity 61.6%; Pred. No. 7.9e-58;  
 Matches 146; Conservative 32; Mismatches 59; Indels 0; Gaps 0;

```

QY 1 LLALVQGSDFLLAKSLDIAPGDMIDKAYRRLSLIKHPECSTMSGGKILIEHLAKOENRF 60
Db 164 LLALAKGIDFLLIQQTLDEAAGDTLIDKARRLSLNHHECGTLLSGQALIERLAKESDRL 223
QY 61 HFIDKPIPLHAHAKCDFSPFGLQHVYDULKIMKKEKKEGIRKQTLSSNADIATVQHTMAC 120
Db 224 AFHFIISPGQVYDNCNFSFAGLRQTQTLGAIKKKEKEGVEAGQLSCVKDIIAASQHTVAS 283
QY 121 HLVKRTHRALIFCKQRDLLPQNNNAVLVASGVASNFYIRRLAEILTNATQCTLLCPPEPL 180
Db 284 HIAKRTHEALIFCKSGKGLPEQNPRTLIVSGVASNEYIRQIOLKITDATTGHLICPEPSKF 343
QY 181 CTNDGIMTAMNGIIEBLRAGLGIHLHDEGIRYEPKCPGLVDISKVEEASTKYVQLMK 237
Db 344 CTNDGIMTAMNGIERIKQGGGLISSEEVSYEPKAPLAGLDTIVSEVBAALIKVPELKL 400

```

## RESULT 9

AC	Q8JBF7	PRELIMINARY;	PRT;	404 AA
DT	01-OCT-2002 (TREMBLrel. 22. Created)			
DT	01-OCT-2002 (TREMBLrel. 22. Last sequence update)			
DT	01-MAR-2004 (TREMBLrel. 26. Last annotation update)			
DE	SI:d727B14.6 (Novel glycoprotease).			
GN	Name=SI:d727B14.6;			
OS	Brachydanio rerio (Zebrafish) (Danio rerio).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;			
OC	Cyprinidae; Danio.			
OX	NCBI_TaxID=7955;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RA	Clark G.;			
RL	Submitted (JUN-2003) to the EMBL/Genbank/DBJ databases.			
DR	EMBL, AL672217; CAD43443.1; ..			
DR	MEROPS M22.004; ..			
DR	GO: GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.			
DR	GO: GO:0008233; F:peptidase activity; IEA.			
DR	GO: GO:0008270; F:zinc ion binding; IEA.			
DR	GO: GO:0005608; P:proteolysis and peptidolysis; IEA.			
DR	InterPro: IPR000905; Peptidase M22.			
DR	InterPro: IPR009180; Pept_M22_Ostalg1.			
DR	Pfam: PF00814; Peptidase_M22; 1.			
DR	PIRSF: PIRSF004537; Ostalgic_dpts; 1.			
DR	PRINTS: PR00789; OSIALOPTAB.			
DR	ProDom: PD002367; Peptidase_M22; 1.			
DR	TIGRFAMS: TIGR00329; gcp; 1.			
KW	PROTEASE.			
SQ	SEQUENCE 404 AA; 44027 MW; 6FE98653A651860F CRC64;			

Query Match	60.2%;	Score 747;	DB 2;	Length 404;
Best Local Similarity	61.2%;	Pred. No. 1.8e-57;		
Matches 145;	Conservative 33;	Mismatches 59;	Indels 0;	Gaps 0;

[illegible]

## RESULT 10

070918			
ID	070918	PRELIMINARY;	PRT; 401 AA.
AC	070918:		
DT	01-MAR-2004 (TrEMBLrel. 26, Created)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	AGP14990 (Fragment).		
OS	Name=agCG46164; ORFNames=ENSANG00000007922;		
GN	Anopheles gambiae str. PEST.		
CC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygnota;		
CC	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.		
OK	NCBI TaxID=180454;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PEST;		
CC	Anopheles Genome Sequencing Consortium;		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
CC	-1- CAUTION: The sequence shown here is derived from an		
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; AAAB01008900; EAA09387.1; -		
DR	GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.		
DR	GO; GO:0008270; F:Zinc ion binding; IEA.		
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR000905; Peptidase_M22		
DR	InterPro; IPR009180; Pept_M22_Ostia1.		
DR	Pfam; PF00814; Peptidase_M22; 1.		
DR	PIRSF; PIRSF004537; Ostia1c_ptds; 1.		
DR	PRINTS; PR00789; OstiaLOPTASE.		
DR	Prodom; PD002367; Peptidase_M22; 1.		
DR	TIGRFAMs; TIGR00329; gcp; 1.		
FT	NON TER		
SE	SEQUENCE	401 AA; 43849 MW; CC9426723D3FD4F1 CRC64;	

Query Match 29.2%; Score 361.5; DB 2; Length 401;  
Best Local Similarity 36.3%; Pred. No. 1.9e-23;  
Matches 90; Conservative 44; Mismatches 103; Indels 11; Gaps 4;

QY	1	LLALVQGSDFLLQKSLDIAFGMDLKVARLSLILKHPEGSTMSGKAIENHLAKOG--	57
Db	147	LLVFEVSTARPRLLGETLDDAAGEALDKIARLKL RNVAKYAQMSGGQAI EAAVQQAQAK	206
QY	58	--NRFEDIKRPLHAAKNCDSFTGLQVHTDKIMKGEABEGIEKQIILSSAADIAATVQ	115
Db	207	DTSAYEPPL--PLSKRDCOFSPAGIKVATATHILIERESTLHAPDALIPVEAFCAQFL	264
QY	116	HTMAHLVARTRRATLFCFKORDLT--PQNNATLVASGVAASNYIPRALEILTNAQCTL	173
Db	265	KGVTRHMLHRTORALEYCERRLLPBDABEHRSLVVSQVACNDVITNALSSMAAQGYST	324
QY	174	LCPRRLCTDNGIMIAMNGIERLRA--GLGIHLIDIEGIRYBPXCPYGVDISKEVGEASIK	231
Db	325	YRPFKCLCTDNGTMIAMNGMERKLLAKDPYAEMLTKYEQVUIDISGRCPIGDSLIDDVKEANIA	384
QY	232	VPQLKMEI	239
Db	385	CKMAKVDI	392

## RESULT 11

ID	PRELIMINARY;	PRT;	323 AA.
Q960S6			
Q960S6			
DT	01-DEC-2001	(TREMBLrel. 19,	Created)
DT	01-DEC-2001	(TREMBLrel. 19,	last sequence update)
DT	01-MAR-2004	(TREMBLrel. 26,	last annotation update)
DE	LD37221p.		
GN	ORFNames=CG14231;		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxId=7227;		

[1]  
RA SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,  
RA Chame M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,  
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY051882; AAK93306.1; -  
DR FlyBase: FBgn0031060; CG14231.  
DR GO: GO:0008450; F:O:staloglycoprotein endopeptidase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR000905; Peptidase M22.  
DR InterPro: IPR009180; Pept M22\_Ostalg1.  
DR Pfam: PF00814; Peptidase M22; 1.  
DR PIRSF: PIRSF004537; Ostalg1c\_ptids; 1.  
DR PRINTS: PR00789; OSIALOPTASE.  
SQ SEQUENCE 323 AA; 35828 MW; BEDB54D9A2BF35A CRC64;  
  
Query Match 26.8%; Score 332.5; DB 2; Length 323;  
Best Local Similarity 35.9%; Pred. No. 5.3e-21;  
Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;  
  
QY 2 LALVGVSDFLILGKSLDIAPGMDLVKVARSLIKPECSGSGKAIEMHLAK-QGNRF 60  
DB 81 LVVANGGRGLTLGCTDDADGEAFDKIGRLRLHLPEYLMNGRAIEHNAQLASPL 140  
QY 61 HEDIKPEPLHAKNCDFSGFTQVHTDKIMKEKEGIEKQILLSSADIAATVQHTAC 120  
DB 141 AVEFPEPLAQQRCNCFSPAGIKNNNSFRAIRARERAPRPPGVISNYGDFCAGLLRSVR 200  
QY 121 HLYKRTALFLFC--KQDRLPQNNNAVLAGGVASNYTIRALBITLNAQTCLICPP 178  
DB 201 HLMKRTQRALEYCLLPHQGLGDPPTPLVMSGVANDAIYANIETHLAQYGRSFRPSK 260  
QY 179 RLCTDNGIMIANNGIERIAGLILHIDE-GIRYEPKPCLVDSKEVGEA 228  
DB 261 RYSDNGVMIAHMGVEQL-----LQDKKASTRTDYD--SIDIQSGAGFA 302  
  
RESULT 12  
QYVMD6 PRELIMINARY; PRT; 409 AA.  
AC QYVMD6; 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE CG14231-PA.  
GN ORFNames=CG14231;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RA SEQUENCE FROM N.A.  
RX MEDLINE=20196006; Pubmed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Byrne C.A., Gocayne J.D.,  
RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers J.H., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
RA Bailew R.M., Beasly A., Bendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolintiner P.,  
RA Borokhov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Burks J.C., Bustam D.A., Butler H., Cauden L.B., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,  
RA Jalali M., Kalush P., Kapen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
RA Kismel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
RA Palazon D.M., Pittman G.S., Pan S., Pollard D., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupki M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Gibbs R.A., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of Drosophila melanogaster.";  
Science 287:2185-2195(2000).  
[2]  
RA SEQUENCE FROM N.A.  
RX MEDLINE=22426065; Pubmed=12537568;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverly T., Muny D.M., Nelson C.R.,  
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodegrem E.J.,  
RA Svitskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
RA "Finishing a whole-genome shotgun: Release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence.";  
RT Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
[3]  
RA SEQUENCE FROM N.A.  
RX MEDLINE=22426070; Pubmed=12537573;  
RA Kaminler J.S., Bergman C.M., Krommiller B., Carlson J., Svitskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.,  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomic perspective.";  
RT Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
[4]  
RA SEQUENCE FROM N.A.  
RX MEDLINE=22426069; Pubmed=12537572;  
RA Miya S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminler J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Belencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.,  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RT Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
[5]  
RA SEQUENCE FROM N.A.  
RX Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
[6]  
RA SEQUENCE FROM N.A.  
RX Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
EMBL: AB003513; AAF49008.1; -  
Intact; QYVMD6; -  
FlyBase: FBgn0031060; CG14231.  
DR GO: GO:0008450; F:O:staloglycoprotein endopeptidase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR000905; Peptidase\_M22.

DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 SQ SEQUENCE 409 AA; 45328 MW; 979766E7D15538 CRC64;

Query Match 26.8%; Score 332.5; DB 2; Length 409;  
 Best Local Similarity 35.9%; Pred. No. 6.9e-21;  
 Matches 83; Conservative 38; Mismatches 97; Indels 13; Gaps 5;

2 LALVQGVSDFLLEKSDIAPGDMIDKVARRLSLIKHECSTSGSKAIEHLAK-QGNRF 60  
 167 LVVANGGRLLTGLGTDIDDAFGAEPDKIGRLRIHLPEYRLMNGRAIEHAQIASDPL 226  
 61 HFDIKPPLHAAKCDPFGTQVHTDKIMKEKEEGIEKQQLSSADIAATVQHTMAC 120  
 227 AVEFPLPLAQRNCFSPAGIKNNNSFRAIRARERPRPGVISNYGDFCAGLLRSYSR 286  
 121 HLVKRTIRALLFC-KORDLLPQNNNAVLAGGVAASNYIRALIELTNATQCTLLCPP 178  
 287 HLMKRTQRALEYCLPHRQLFGDTPPTLVMSGVANNDAIYANIEHLAAQYGCGRFSR 346  
 179 RLCTDNGIMIAMNGIERLRAGLGIHLIDIE-GIRYEPKPCPLGDISKEVGEA 228  
 347 RYCSDNQVMIAHMGVEOL-----LQDKKASTRIDYD--SIDIQSGAGFA 388

RESULT 13  
 022145 PRELIMINARY; PRT; 480 AA.

01-JUN-1998 (TREMBlrel. 05, Created)  
 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 Putative O-sialoglycoprotein endopeptidase (Sialoglycoprotease GCP1).  
 Name:At2g45270; Synonyms=GCP1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC euroids II; Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_TaxID=3702;

[1]  
 SEQUENCE FROM N.A.  
 Rounseley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,  
 Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,  
 Venier J.C.;  
 Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[2]  
 SEQUENCE FROM N.A.  
 Town C.D., Kaul S.;  
 Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

[3]  
 SEQUENCE FROM N.A.  
 Hausenhi K., Adamka I.;  
 Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

[4]  
 SEQUENCE FROM N.A.  
 Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bower L.,  
 Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 Seki M., Shimizu P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,  
 Theologis A.;  
 Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

[5]  
 SEQUENCE FROM N.A.  
 Yamada K., Banh J., Chan M.M., Chang C.H., Chang B., Dale J.M.,  
 Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bower L.,

RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 Seki M., Shimizu P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

RA Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC002387; AAB82636.2; -;  
 DR EMBL; AY024338; AAB00530.1; -;  
 DR EMBL; AY063864; AAL36220.1; -;  
 DR EMBL; AY117283; AAM51358.1; -;  
 DR PIR; E84888; E84888.  
 DR GO; GO:0008450; P:O-sialoglycoprotein endopeptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR00905; Peptidase\_M22.  
 DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
 DR Pfam; PF00814; Peptidase\_M22; 1.  
 DR PIRSF; PIRSF004537; Osialglc\_ptds; 1.  
 DR PRINTS; PR00789; OSIALOPTASE.  
 DR ProDom; PD002367; Peptidase\_M22; 1.  
 DR TIGRFAMs; TIGR00329; gcp; 1.  
 KW Protease.  
 SQ SEQUENCE 480 AA; 52995 MW; 20DD6A6ACCIFFPAD CRC64;

Query Match 26.1%; Score 324; DB 2; Length 480;  
 Best Local Similarity 34.1%; Pred. No. 4.7e-20;  
 Matches 86; Conservative 34; Mismatches 86; Indels 46; Gaps 6;

1 LALVQGVSDFLLEKSDIAPGDMIDKVARRLSLIKHECSTSGSKAIEHLAKQGNRF 60  
 224 LVVANGGRLLTGLGTDIDDAFGAEPDKIGRLRIHLPEYRLMNGRAIEHAQIASDPL 278  
 61 HFDIKPPLHAAKCDPFGTQVHTDKIMKEKEEGIEKQQLSSA-----ADIAA 112  
 279 SVKFAVPMKHYKDCNFSYAGLTQVRLAIEAKE-----IDAKCVSATSATEDRRNRADIAA 334  
 113 TVQHTMACGLVARTIRALLFCQKORDLLPQNNNAVLAGGVAASNYIRALIELTNATQCT 172  
 335 SFQRAVALHLERKCEBALDWALE--LEPSIKHMTVSGGVAASNYIRALIELTNATQCT 391  
 173 LCCPPRLCTDNGIMIAMNGIERLRAGLGIHLIDIEGIRY-----PKCP 216  
 392 LVCPPLSLCTDNGVAVATGLHFRVG-----RIDPPPAIEPBDYVDLARPWP 441

01-JUN-1998 (TREMBlrel. 28, Created)  
 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 O-sialoglycoprotein endopeptidase (Fragment).  
 OS Hyacinthus orientalis.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Hyacinthaceae;  
 OC Hyacinthus.  
 NCBI\_TaxID=82025;  
 [1]  
 SEQUENCE FROM N.A.  
 TISSUE=Floral meristem 5-10 days when regenerated in vitro;  
 Fan J.H., Ma Y., Zhang X.S.;  
 "Hyacinthus orientalis O-sialoglycoprotein endopeptidase like protein  
 mRNA, expressed during regeneration of floral bud."  
 Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
 EMBL; AY389554; AAT08652.1; -;  
 InterPro; IPR00905; Peptidase\_M22.  
 PRINTS; PR00789; OSIALOPTASE.

DR Prodom; PD002367; Peptidase\_M22; 1.  
FT NON TER 1  
FT NON TER 255  
SQ SEQUENCE 255 AA; 28516 MW; 529D22666C284106 CRC64;

Query Match 24.2%; Score 299.5; DB 2; Length 255;  
Best Local Similarity 37.4%; Pred. No. 3.3e-18;  
Matches 79; Conservative 26; Mismatches 79; Indels 27; Gaps 5;

QY 14 LGSGLDIARGMLDVARRLSLIKHPECGTSGKAIITHLAKQGRFHDIPRLHAKN 73  
DB 1 LGTIDDDIGAYDPTARMLGL-----DMGKGGGALBELALEGANSVKFKI PWRQHKD 55  
QY 74 CDFSGFTGLQHTYDKTIIMKEKEKEGKQILSSA-----ADIAATVQHTMACHLVKRT 126  
DB 56 CNFESYAGIK--TQRLALAESRNICIDKSPISLATCHDRMRADLAASFGVAVLHLEDRG 113  
QY 127 HRAILFCCKQKRDLLPQNNAVLVASGGVANSFYIRRALILTNATQCTLLCPPRLCTDNGI 186  
DB 114 ARAIEMWALIEIPTVQ---CLVVGSGVANSKRYRSRLNHLVKTGLRLVCPSPSLCTDNGV 170  
QY 187 MIANNGLERLRAAGILHDIEGIRYEPKCP 217  
DB 171 MWAMTGIENFLVG-----RYDPPPPV 191

## RESULT 15

Q92LH8 PRELIMINARY; PRT; 360 AA.  
ID Q92LH8  
AC Q92LH8  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE PROBABLE O-SIALOGLYCOPROTEIN ENDOPEPTIDASE (EC 3.4.24.57).  
GN ORFNames=SMC03230;  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
OX NCBI\_TaxID=382;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Boistard P., Becker A., Boutry M., Cadieu B., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
RA Pohl T., Portetleau D., Puehler A., Purnelle B., Rampsberger U.,  
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium meliloti strain 1021.",  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
DR EMBL; AL591792; CAC47657.1; -.  
DR GO; GO:0016787; F:hydrolyase activity; IEA.  
DR GO; GO:0008450; F:O-sialoglycoprotein endopeptidase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR00905; Peptidase\_M22.  
DR InterPro; IPR009180; Pept\_M22\_Osialgl.  
DR Pfam; PF00814; Peptidase\_M22\_1.  
DR PIRSF; PIRSF004537; Osialglc\_ptide; 1.  
DR PRINTS; PR00789; OSIALOPTASE.  
DR Prodom; PD002367; Peptidase\_M22; 1.  
DR TIGRFAMs; TIGR00329; gcp; 1.  
KM Complete proteome; Hydrolyase.  
SQ SEQUENCE 360 AA; 37906 MW; A07F946AB562EA86 CRC64;

Query Match 21.2%; Score 262.5; DB 2; Length 360;  
Best Local Similarity 32.8%; Pred. No. 9e-15;  
Matches 78; Conservative 32; Mismatches 85; Indels 43; Gaps 7;

QY 2 LALVQGVSDPFLLLKSLDIARGMDLKVARRLSLIKHECGTSGKAIETHLAKQGRFH 61  
DB 145 LILVKGVEYERKWTITDDALGEAPDKTKAKLGL-PYP-----GGPAVERAAQGNAR 197

QY 62 PDIKPELHAKKCDFSFTGLQ-----HYTDKIIMKEKEKEGKQILSSAADI 110  
DB 198 FDPPEPLVDADALDFSFSLKTAVRQAQSLGPTTODI-----ADV 239  
QY 111 AATVQHTMACHLVKRTHRAILFCCKQKRDLLPQNNAVLVASGGVANSFYIRRALILTNATQ 170  
DB 240 CASFGRAISRTLRDRVGRGLKRFR-ADFASVDQPALVAVGVAANQTLRRTLQSLCDEHG 298  
QY 171 CTLLCPPRLCTDNGIMIANGLERLRAAGILHDIEGIRYEPKCP--PLGVDISKEVG 226  
DB 299 FFFIAPPIQLCTDNAMIAMMAGERLAAGL---PADGLDAAPRSRWPIDSEKALIG 352

Search completed: February 16, 2005, 13:08:41  
Job time : 41.5261 secs